

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/00, 15/29, A01H 1/00, 5/00, 5/10</b>		<b>A1</b>	(11) International Publication Number: <b>WO 98/07842</b>
			(43) International Publication Date: 26 February 1998 (26.02.98)
(21) International Application Number: PCT/US97/14659			(74) Agents: BASTIAN, Kevin, L. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111 (US).
(22) International Filing Date: 19 August 1997 (19.08.97)			
(30) Priority Data: 08/700,152 20 August 1996 (20.08.96) US 08/879,827 20 June 1997 (20.06.97) US			
(60) Parent Application or Grant (63) Related by Continuation US Not furnished (CIP) Filed on Not furnished			
(71) Applicant (for all designated States except US): THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 22nd floor, 300 Lakeside Drive, Oakland, CA 94612 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): JOFUKU, K., Diane [US/US]; 475 Sims Road, Santa Cruz, CA 95060 (US). OKAMURO, Jack, K. [US/US]; 475 Sims Road, Santa Cruz, CA 95060 (US).			(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
<b>Published</b> With international search report.			
(54) Title: METHODS FOR IMPROVING SEEDS			
(57) Abstract  The invention provides methods of modulating seed mass and other traits in plants. The methods involve producing transgenic plants comprising a recombinant expression cassette containing an <i>ADC</i> nucleic acid linked to a plant promoter.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## METHODS FOR IMPROVING SEEDS

## REFERENCE TO RELATED APPLICATIONS

5 This is a continuation in part of copending application USSN 08/879,827  
(Attorney Docket No. 02307O-067210, filed June 20, 1997, which is a continuation in  
part of USSN 08/700,152, filed August 20, 1996, both of which are incorporated herein  
by reference.

## FIELD OF THE INVENTION

10 The present invention is directed to plant genetic engineering. In particular,  
it relates to new methods for modulating mass and other properties of plant seeds.

## BACKGROUND OF THE INVENTION

15 The pattern of flower development is controlled by the floral meristem, a  
complex tissue whose cells give rise to the different organ systems of the flower. Genetic  
and molecular studies have defined an evolutionarily conserved network of genes that  
control floral meristem identity and floral organ development in *Arabidopsis*, snapdragon,  
and other plant species (*see, e.g.*, Coen and Carpenter, *Plant Cell* 5:1175-1181 (1993) and  
Okamuro *et al.*, *Plant Cell* 5:1183-1193 (1993)). In *Arabidopsis*, a floral homeotic gene  
20 *APETALA2* (*AP2*) controls three critical aspects of flower ontogeny - the establishment of  
the floral meristem (Irish and Sussex, *Plant Cell* 2:741-753 (1990); Huala and Sussex,  
*Plant Cell* 4:901-913 (1992); Bowman *et al.*, *Development* 119:721-743 (1993); Schultz  
and Haughn, *Development* 119:745-765 (1993); Shannon and Meeks-Wagner, *Plant Cell*  
5:639-655 (1993)), the specification of floral organ identity (Komaki *et al.*, *Development*  
25 104:195-203 (1988)); Bowman *et al.*, *Plant Cell* 1:37-52 (1989); Kunst *et al.*, *Plant Cell*  
1:1195-1208 (1989)), and the temporal and spatial regulation of floral homeotic gene  
expression (Bowman *et al.*, *Plant Cell* 3:749-758 (1991); Drews *et al.*, *Cell* 65:91-1002  
(1991)).

30 One early function of *AP2* during flower development is to promote the  
establishment of the floral meristem. *AP2* performs this function in cooperation with at  
least three other floral meristem genes, *APETALA1* (*API*), *LEAFY* (*LFY*), and  
*CAULIFLOWER* (*CAL*) (Irish and Sussex (1990); Bowman, *Flowering Newsletter* 14:7-19

(1992); Huala and Sussex (1992); Bowman *et al.*, (1993); Schultz and Haughn, (1993); Shannon and Meeks-Wagner, (1993)). A second function of *AP2* is to regulate floral organ development. In *Arabidopsis*, the floral meristem produces four concentric rings or whorls of floral organs - sepals, petals, stamens, and carpels. In weak, partial loss-of-function *ap2* mutants, sepals are homeotically transformed into leaves, and petals are transformed into pollen-producing stamenoid organs (Bowman *et al.*, *Development* 112:1-20 (1991)). By contrast, in strong *ap2* mutants, sepals are transformed into ovule-bearing carpels, petal development is suppressed, the number of stamens is reduced, and carpel fusion is often defective (Bowman *et al.*, (1991)). Finally, the effects of *ap2* on floral organ development are in part a result of a third function of *AP2*, which is to directly or indirectly regulate the expression of several flower-specific homeotic regulatory genes (Bowman *et al.*, *Plant Cell* 3:749-758 (1991); Drews *et al.*, *Cell* 65:91-1002 (1991); Jack *et al.* *Cell* 68:683-697 (1992); Mandel *et al.* *Cell* 71: 133-143 (1992)).

Clearly, *Ap2* plays a critical role in the regulation of *Arabidopsis* flower development. Yet, little is known about how it carries out its functions at the cellular and molecular levels. A spatial and combinatorial model has been proposed to explain the role of *AP2* and other floral homeotic genes in the specification of floral organ identity(see, e.g., Coen and Carpenter, *supra*). One central premise of this model is that *AP2* and a second floral homeotic gene *AGAMOUS* (*AG*) are mutually antagonistic genes. That is, *AP2* negatively regulates *AG* gene expression in sepals and petals, and conversely, *AG* negatively regulates *AP2* gene expression in stamens and carpels. In situ hybridization analysis of *AG* gene expression in wild-type and *ap2* mutant flowers has demonstrated that *AP2* is indeed a negative regulator of *AG* expression. However, it is not yet known how *AP2* controls *AG*. Nor is it known how *AG* influences *AP2* gene activity.

The *AP2* gene in *Arabidopsis* has been isolated by T-DNA insertional mutagenesis as described in Jofuku *et al.* *The Plant Cell* 6:1211-1225 (1994). *AP2* encodes a putative nuclear factor that bears no significant similarity to any known fungal, or animal regulatory protein. Evidence provided there indicates that *AP2* gene activity and function are not restricted to developing flowers, suggesting that it may play a broader role in the regulation of *Arabidopsis* development than originally proposed.

In spite of the recent progress in defining the genetic control of plant development, little progress has been reported in the identification and analysis of genes

effecting agronomically important traits such as seed size, protein content, oil content and the like. Characterization of such genes would allow for the genetic engineering of plants with a variety of desirable traits. The present invention addresses these and other needs.

## SUMMARY OF THE INVENTION

The present invention provides methods of modulating seed mass and other traits in plants. The methods involve providing a plant comprising a recombinant expression cassette containing an *ADC* nucleic acid linked to a plant promoter. The plant is either selfed or crossed with a second plant to produce a plurality of seeds. Seeds with the desired trait (e.g., altered mass) are then selected.

In some embodiments, transcription of the *ADC* nucleic acid inhibits expression of an endogenous *ADC* gene or activity the encoded protein. In these embodiments, the step of selecting includes the step of selecting seed with increased mass or another trait. The seed may have, for instance, increased protein content, carbohydrate content, or oil content. In the case of increased oil content, the types of fatty acids may or may not be altered as compared to the parental lines. In these embodiments, the *ADC* nucleic acid may be linked to the plant promoter in the sense or the antisense orientation. Alternatively, expression of the *ADC* nucleic acid may enhance expression of an endogenous *ADC* gene or *ADC* activity and the step of selecting includes the step of selecting seed with decreased mass. This embodiment is particularly useful for producing seedless varieties of crop plants.

If the first plant is crossed with a second plant the two plants may be the same or different species. The plants may be any higher plants, for example, members of the families Brassicaceae or Solanaceae. In making seed of the invention, either the female or the male parent plant can comprise the expression cassette containing the *ADC* nucleic acid. In preferred embodiments, both parents contain the expression cassette.

In the expression cassettes, the plant promoter may be a constitutive promoter, for example, the CaMV 35S promoter. Alternatively, the promoter may be a tissue-specific promoter. Examples of tissue specific expression useful in the invention include fruit-specific, seed-specific (e.g., ovule-specific, embryo-specific, endosperm-specific, integument-specific, or seed coat-specific) expression.

The invention also provides seed produced by the methods described above. The seed of the invention comprise a recombinant expression cassette containing an *ADC* nucleic acid. If the expression cassette is used to inhibit expression of endogenous *ADC* expression, the seed will have a mass at least about 20% greater than the average mass of seeds of the same plant variety which lack the recombinant expression cassette. If the expression cassette is used to enhance expression of *ADC*, the seed will have a mass at least about 20% less than the average mass of seeds of the same plant variety which lack the recombinant expression cassette. Other traits such as protein content, carbohydrate content, and oil content can be altered in the same manner.

### Definitions

The phrase "nucleic acid sequence" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It includes chromosomal DNA, self-replicating plasmids, infectious polymers of DNA or RNA and DNA or RNA that performs a primarily structural role.

The term "promoter" refers to a region or sequence determinants located upstream or downstream from the start of transcription and which are involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "plant promoter" is a promoter capable of initiating transcription in plant cells.

The term "plant" includes whole plants, plant organs (e.g., leaves, stems, flowers, roots, etc.), seeds and plant cells and progeny of same. The class of plants which can be used in the method of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous.

A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants. As defined here, a modified *ADC* coding sequence which is heterologous to an operably linked *ADC* promoter

does not include the T-DNA insertional mutants (*e.g.*, *ap2-10*) as described in Jofuku *et al.* *The Plant Cell* 6:1211-1225 (1994).

A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, and the like. Such a plant containing the exogenous nucleic acid is referred to here as an  $R_1$  generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are descendants of such a plant.

An "ADC (AP2 domain containing) nucleic acid" or "ADC polynucleotide sequence" of the invention is a subsequence or full length polynucleotide sequence of a gene which, encodes an polypeptide containing an AP2 domain and when present in a transgenic plant, can be used to modulate seed properties in seed produced by the plant. An exemplary nucleic acid of the invention is the *Arabidopsis* AP2 sequence as disclosed in Jofuku *et al.* *The Plant Cell* 6:1211-1225 (1994). The GenBank accession number for this sequence is U12546. As explained in detail below a family of *RAP2* (related to *AP2*) genes have been identified in *Arabidopsis*. The class of nucleic acids claimed here falls into at least two subclasses (AP2-like and EREBP-like genes), which are distinguished by, for instance, the number of AP2 domains contained within each polypeptide and by sequences within certain conserved regions. The differences between these two subclasses are described in more detail below. ADC polynucleotides are defined by their ability to hybridize under defined conditions to the exemplified nucleic acids or PCR products derived from them. An ADC polynucleotide (*e.g.*, *AP2* or *RAP2*) is typically at least about 30-40 nucleotides to about 3000, usually less than about 5000 nucleotides in length. Usually the nucleic acids are from about 100 to about 2000 nucleotides, often from about 500 to about 1700 nucleotides in length.

ADC nucleic acids, as explained in more detail below, are a new class of plant regulatory genes that encode ADC polypeptides, which are distinguished by the presence of one or more of a 56-68 amino acid repeated motif, referred to here as the "AP2 domain". The amino acid sequence of an exemplary AP2 polypeptide is shown in Jofuku *et al.*, *supra*. One of skill will recognize that in light of the present disclosure various modifications (*e.g.*, substitutions, additions, and deletions) can be made to the

sequences shown there without substantially affecting its function. These variations are specifically covered by the terms ADC polypeptide or *ADC* polynucleotide.

In the case of both expression of transgenes and inhibition of endogenous genes (e.g., by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only "substantially identical" to a sequence of the gene from which it was derived. As explained below, these substantially identical variants are specifically covered by the term *ADC* nucleic acid.

In the case where the inserted polynucleotide sequence is transcribed and translated to produce a functional polypeptide, one of skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same polypeptide. These variants are specifically covered by the terms "*ADC* nucleic acid", "*AP2* nucleic acid" and "*RAP2* nucleic acid". In addition, the term specifically includes those full length sequences substantially identical (determined as described below) with an *ADC* polynucleotide sequence and that encode proteins that retain the function of the *ADC* polypeptide (e.g., resulting from conservative substitutions of amino acids in the *AP2* polypeptide). In addition, variants can be those that encode dominant negative mutants as described below.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The term "complementary to" is used herein to mean that the complementary sequence is identical to all or a portion of a reference polynucleotide sequence.

Sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a "comparison window" to identify and compare local regions of sequence similarity. A "comparison window", as used herein, refers to a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch *J. Mol. Biol.* 48:443 (1970), by



the search for similarity method of Pearson and Lipman *Proc. Natl. Acad. Sci. (U.S.A.)* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

5 "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by  
10 determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

The term "substantial identity" of polynucleotide sequences means that a  
15 polynucleotide comprises a sequence that has at least 60% sequence identity, preferably at least 80%, more preferably at least 90% and most preferably at least 95%, compared to a reference sequence using the programs described above (preferably BLAST) using standard parameters. One of skill will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide  
20 sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 35%, preferably at least 60%, more preferably at least 90%, and most preferably at least 95%. Polypeptides which are "substantially similar" share sequences as noted above except that residue positions which  
25 are not identical may differ by conservative amino acid changes. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is  
30 asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side

chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other, or a third nucleic acid, under stringent conditions. Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5° C lower than the thermal melting point (T<sub>m</sub>) for the specific sequence at a defined ionic strength and pH. The T<sub>m</sub> is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least about 60°C.

In the present invention, genomic DNA or cDNA comprising *ADC* nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic acid sequences disclosed here. For the purposes of this disclosure, stringent conditions for such hybridizations are those which include at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C, for 20 minutes, or equivalent conditions. Other means by which nucleic acids of the invention can be identified are described in more detail below.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A shows amino acid sequence alignment between AP2 direct repeats AP2-R1 (aa 129-195) and AP2-R2 (aa 221-288). Solid and dashed lines between the two sequences indicate residue identity and similarity, respectively. Arrows indicate the positions of the ap2-1, ap2-5, and ap2-10 mutations described in Jofuku *et al.* (1994). The bracket above the AP2-R1 and AP2-R2 sequences indicates the residues capable of forming amphipathic  $\alpha$ -helices shown in Figure 1B.

Figure 1B is a schematic diagram of the putative AP2-R1 (R1) and AP2-R2 (R2) amphipathic  $\alpha$ -helices. The NH<sub>2</sub> terminal ends of the R1 and R2 helices begin at residues Phe-160 and Phe-253 and rotate clockwise by 100° per residue through Phe-177 and Cys-270, respectively. Arrows directed toward or away from the center of the helical

wheel diagrams indicate the negative or positive degree of hydrophobicity as defined by Jones *et al. J. Lipid Res.* 33: 87-296 (1992).

Figure 2 shows an antisense construct of the invention. pPW14.4 (which is identical to pPW15) represents the 13.41 kb AP2 antisense gene construct used in plant transformation described here. pPW14.4 is comprised of the AP2 gene coding region in a transcriptional fusion with the cauliflower mosaic virus 35S (P35S) constitutive promoter in an antisense orientation. The Ti plasmid vector used is a modified version of the pGSJ780A vector (Plant Genetic Systems, Gent, Belgium) in which a unique EcoR1 restriction site was introduced into the BamH1 site using a ClaI-EcoR1-BamH1 adaptor. The modified pGSJ780A vector DNA was linearized with EcoR1 and the AP2 coding region inserted as a 1.68 kb EcoR1 DNA fragment from AP2 cDNA plasmid cAP2#1 (Jofuku *et al.*, 1994) in an antisense orientation with respect to the 35S promoter. KmR represents the plant selectable marker gene NPTII which confers resistance to the antibiotic kanamycin to transformed plant cells carrying an integrated 35S-AP2 antisense gene. Boxes 1 and 5 represent the T-DNA left and right border sequences, respectively, that are required for transfer of T-DNA containing the 35S-AP2 antisense gene construct into the plant genome. Regions 2 and 3 contain T-DNA sequences. Box 3 designates the 3' octopine synthase gene sequences that function in transcriptional termination. Region 6 designates bacterial DNA sequences that function as a bacterial origin of replication in both *E. coli* and *Agrobacterium tumefaciens*, thus allowing pPW14.4 plasmid replication and retention in both bacteria. Box 7 represents the bacterial selectable marker gene that confers resistance to the antibiotics streptomycin and spectinomycin and allows for selection of *Agrobacterium* strains that carry the pPW14.4 recombinant plasmid.

Figure 3 shows a sense construct of the invention. pPW12.4 (which is identical to pPW9) represents the 13.41 kb AP2 sense gene construct used in plant transformation described here. pPW12.4 is comprised of the AP2 gene coding region in a transcriptional fusion with the cauliflower mosaic virus 35S (P35S) constitutive promoter in a sense orientation. The Ti plasmid vector used is a modified version of the pGSJ780A vector (Plant Genetic Systems, Gent, Belgium) in which a unique EcoR1 restriction site was introduced into the BamH1 site using a ClaI-EcoR1-BamH1 adaptor. The modified pGSJ780A vector DNA was linearized with EcoR1 and the AP2 coding region inserted as a 1.68 kb EcoR1 DNA fragment from AP2 cDNA plasmid cAP2#1 (Jofuku *et al.*, 1994)

in a sense orientation with respect to the 35S promoter. KmR represents the plant selectable marker gene NPTII which confers resistance to the antibiotic kanamycin to transformed plant cells carrying an integrated 35S-AP2 antisense gene. Boxes 1 and 5 represent the T-DNA left and right border sequences, respectively, that are required for transfer of T-DNA containing the 35S-AP2 sense gene construct into the plant genome. Regions 2 and 3 contain T-DNA sequences. Box 3 designates the 3' octopine synthase gene sequences that function in transcriptional termination. Region 6 designates bacterial DNA sequences that function as a bacterial origin of replication in both *E. coli* and *Agrobacterium tumefaciens*, thus allowing pPW12.4 plasmid replication and retention in both bacteria. Box 7 represents the bacterial selectable marker gene that confers resistance to the antibiotics streptomycin and spectinomycin and allows for selection of *Agrobacterium* strains that carry the pPW12.4 recombinant plasmid.

Figures 4A and 4B show AP2 domain sequence and structure. The number of amino acid residues within each AP2 domain is shown to the right. Sequence gaps were introduced to maximize sequence alignments. The position of amino acid residues and sequence gaps within the AP2 domain alignments are numbered 1-77 for reference. The location of the conserved YRG and RAYD elements are indicated by brackets. Shaded boxes highlight regions of sequence similarity. Positively charged amino acids within the YRG element are indicated by + signs above the residues. The location of the 18-amino acid core region that is predicted to form an amphipathic  $\alpha$ -helix in AP2 is indicated by a bracket. Residues within the RAYD element of each AP2 domain that are predicted to form an amphipathic  $\alpha$ -helix are underlined. Figure 4A shows members of the AP2-like subclass. Amino acid sequence alignment between the AP2 domain repeats R1 and R2 contained within AP2, ANT and RAP2.7 is shown. Brackets above the sequences designate the conserved YRG and RAYD blocks described above. The filled circle and asterisk indicate the positions of the *ap2-1*, and *ap2-5* mutations, respectively. Amino acid residues that constitute a consensus AP2 domain motif for AP2, ANT, and RAP2.7 is shown below the alignment with invariant residues shown capitalized. Figure 4B shows members of the EREBP-like subclass. Amino acid sequence alignment between the AP2 domains contained within the tobacco EREBPs and the *Arabidopsis* EREBP-like RAP2 proteins is shown. GenBank accession numbers for EREBP-1, EREBP-2, EREBP-3, and EREBP-4 are D38123, D38126, D38124, and D38125, respectively.

Figure 4C provides schematic diagrams of the putative RAP2.7-R1, AP2-R1, and ANT-R1 amphipathic  $\alpha$ -helices. Amino acid residues within the RAP2.7-R1, AP2-R1, and ANT-R1 motifs shown underlined in A that are predicted to form amphipathic  $\alpha$ -helices are schematically displayed with residues rotating clockwise by 100° per residue to form helical structures. Arrows directed toward or away from the center of the helical wheel diagrams indicate the negative or positive degree of hydrophobicity as defined by Jones *et al. J. Lipid Res.* 33:287-296 (1992). Positively and negatively charged amino acid residues are designated by + and - signs, respectively.

Figure 4D shows schematic diagrams of the putative RAP2.2, RAP2.5, RAP2-12, and EREBP-3 amphipathic  $\alpha$ -helices. Amino acid residues within the RAP2.2, RAP2.5, RAP2-12, and EREBP-3 motifs shown underlined in Figure 4B that are predicted to form amphipathic  $\alpha$ -helices are schematically displayed as described in Figure 4C.

Figure 4C shows sequence alignment between the 25-26 amino acid linker regions in AP2, ANT, and RAP2.7. R1 and R2 designate the positions of the R1 and R2 repeats within AP2, ANT, and RAP2.7 relative to the linker region sequences. Boxes designate invariant residues within the conserved linker regions. Amino acid residues that constitute a consensus linker region motif for AP2, ANT, and RAP2.7 are shown below the alignment with invariant residues shown capitalized. The arrowhead indicates the position of the *ant-3* mutation described by Klucher *et al. Plant Cell* 8:137-153 (1996).

Figure 5 is a schematic diagram of pAP2, which can be used to construct expression vectors of the invention.

Figure 6 is a schematic diagram of pBEL1, which can be used to construct expression vectors of the invention.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

This invention relates to plant *ADC* genes, such as the *AP2* and *RAP2* genes of *Arabidopsis*. The invention provides molecular strategies for controlling seed size and total seed protein using *ADC* overexpression and antisense gene constructs. In particular, transgenic plants containing antisense constructs have dramatically increased seed mass, seed protein, or seed oil. Alternatively, overexpression of *ADC* using a constructs of the invention leads to reduced seed size and total seed protein. Together, data presented here

demonstrate that a number of agronomically important traits including seed mass, total seed protein, and oil content, can be controlled in species of agricultural importance.

#### Isolation of *ADC* nucleic acids

5                   Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's  
10                   specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989).

                  The isolation of *ADC* nucleic acids may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here  
15                   can be used to identify the desired gene in a cDNA or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ, such as flowers, and a cDNA library  
20                   which contains the *ADC* gene transcript is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which *ADC* genes or homologs are expressed.

                  The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned *ADC* gene disclosed here. Probes may be used to hybridize  
25                   with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Alternatively, antibodies raised against an *ADC* polypeptide can be used to screen an mRNA expression library.

                  Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction  
30                   (PCR) technology can be used to amplify the sequences of the *ADC* genes directly from genomic DNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid

sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes.

Appropriate primers and probes for identifying *ADC* sequences from plant tissues are generated from comparisons of the sequences provided in Jofuku *et al.*, *supra*. For a general overview of PCR see *PCR Protocols: A Guide to Methods and Applications*. (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), *Academic Press*, San Diego (1990).

As noted above, the nucleic acids of the invention are characterized by the presence of sequence encoding a AP2 domain. Thus, these nucleic acids can be identified by their ability to specifically hybridize to sequences encoding AP2 domain disclosed here. Primers which specifically amplify AP2 domains of the exemplified genes are particularly useful for identification of particular *ADC* polynucleotides. Primers suitable for this purpose based on the sequence of RAP2 genes disclosed here are as follows:

Name	GenBank Number	Primers
AP2	U12546	JOAP2U 5'-GTTGCCGCTGCCGTAGTG-3' JOAP2L 5'-GGTTCATCCTGAGCCGCATATC-3'
RAP2.1	AF003094	JORAP2.1U 5'-CTCAAGAAGAAGTGCCTAACCACG-3' JORAP2.1L 5'-GCAGAAGCTAGAAGAGCGTCGA-3'
RAP2.2	AF003095	JORAP2.2U 5'-GGAAAATGGGCTGCGGAG-3' JORAP2.2L 5'-GTTACCTCCAGCATCGAACGAG-3'
RAP2.4	AF003097	JORAP2.4U 5'-GCTGGATCTTGTTTCGCTTACG-3' JORAP2.4L 5'-GCTTCAAGCTTAGCGTCGACTG-3'
RAP2.5	AF003098	JORAP2.5U 5'-AGATGGGCTTGAAACCCGAC-3' JORAP2.5L 5'-CTGGCTAGGGCTACGCGC-3'
RAP2.6	AF003099	JORAP2.6U 5'-TTCTTTGCCTCCTCAACCATTG-3' JORAP2.6L 5'-TCTGAGTTCCAACATTTTCGGG-3'
RAP2.7	AF003100	JORAP2.7U 5'-GAAATTGGTAACTCCGGTTCCG-3' JORAP2.7L 5'-CCATTTTGCTTTGGCGCATTAC-3'

RAP2.8	AF003101	JORAP2.8U 5'-GGCGTTACGCCTCTACCGG-3' JORAP2.8L 5'-CGCCGTCTTCCAGAACGTTC-3'
RAP2.9	AF003102	JORAP2.9U 5'-ATCACGGATCTGGCTTGGTTC-3' JORAP2.9L 5'-GCCTTCTTCCGTATCAACGTCG-3'
RAP2.10	AF003103	JORAP2.10U 5'-GTCAACTCCGGCGGTACG-3' JORAP2.10L 5'-TCTCCTTATATACGCCGCCGA-3'
RAP2.11	AF003104	JORAP2.11U 5'-GAGAAGAGCAAAGGCAACAAGAC-3' JORAP2.11L 5'-AGTTGTTAGGAAAATGGTTTGCG-3'
RAP2.12	AF003105	JORAP2.12U 5'-AAACCATTCGTTTTCACTTCGACTC-3' JORAP2.12L 5'-TCACAGAGCGTTTCTGAGAATTAGC-3'

The PCR primers are used under standard PCR conditions (described for instance in Innis *et al.*) using the nucleic acids as identified in the above GenBank accessions as a template. The PCR products generated by any of the reactions can then be used to identify nucleic acids of the invention (*e.g.*, from a cDNA library) by their ability to hybridize to these products. Particularly preferred hybridization conditions use a Hybridization Buffer consisting of: 0.25M Phosphate Buffer (pH 7.2), 1 mM EDTA, 1% Bovine Serum Albumin, 7% SDS. Hybridizations then followed by a first wash with 2.0XSSC + 0.1% SDS or 0.39M Na<sup>+</sup> and subsequent washes with 0.2XSSC + 0.1% SDS or 0.042M Na<sup>+</sup>. Hybridization temperature will be from about 45°C to about 78°C, usually from about 50°C to about 70°C. Followed by washes at 18°C.

Particularly preferred hybridization conditions are as follows:

Hybridization Temp.	Hybrid. Time	Wash Buffer A	Wash Buffer B
78 degrees C	48 hrs	18 degrees C	18 degrees C
70 degrees C	48 hrs	18 degrees C	18 degrees C
65 degrees C	48 hrs	18 degrees C	18 degrees C
60 degrees C	72 hrs	18 degrees C	18 degrees C
55 degrees C	96 hrs	18 degrees C	18 degrees C
45 degrees C	200 hrs	18 degrees C	No wash

If desired, primers that amplify regions more specific to particular ADC genes can be used. The PCR products produced by these primers can be used in the hybridization conditions described above to isolate nucleic acids of the invention.



Name	GenBank Number	Primers
AP2	U12546	AP2U 5'-ATGTGGGATCTAAACGACGCAC-3' AP2L 5'-GATCTTGGTCCACGCCGAC-3'
RAP2.1	AF003094	RAP2.1U 5'-AAG AGG ACC ATC TCT CAG-3' RAP2.1L 5'-AAC ACT CGC TAG CTT CTC-3'
RAP2.2	AF003095	RAP2.2U 5'-TGG TTC AGC AGC CAA CAC-3' RAP2.2L 5'-CAA TGC ATA GAG CTT GAG G-3'
RAP2.4	AF003097	RAP2.4U 5'-ACG GAT TTC ACA TCG GAG-3' RAP2.4L 5'-CTA AGC TAG AAT CGA ATC C-3'
RAP2.5	AF003098	RAP2.5U 5'-TACCGGTTTCGCGCGTAG-3' RAP2.5L 5'-CACCTTCGAAATCAACGACCG-3'
RAP2.6	AF003099	RAP2.6U 5'-TTCCCCGAAAATGTTGGAAGTC-3' RAP2.6L 5'-TGCGAGAGAAAAAATTGGTAGATCG-3'
RAP2.7	AF003100	RAP2.7U 5'- CGA TGG AGA CGA AGA CTC-3' RAP2.7L 5'- GTC GGA ACC GGA GTT ACC-3'
RAP2.8	AF003101	RAP2.8U 5'-TCA CTC AAA GGC CGA GAT C-3' RAP2.8L 5'-TAA CAA CAT CAC CGG CTC G-3'
RAP2.9	AF003102	RAP2.9U 5'-GTG AAG GCT TAG GAG GAG-3' RAP2.9L 5'-TGC CTC ATA TGA GTC AGA G-3'
RAP2.10	AF003103	RAP2.10U 5'-TCCCGGAGCTTTTAGCCG-3' RAP2.10L 5'-CAACCCGTTCCAACGATCC-3'
RAP2.11	AF003104	RAP2.11U 5'-TTCTTCACCAGAAGCAGAGCATG-3' RAP2.11L 5'-CTCCATTTCATTGCATATAGGGACG-3'
RAP2.12	AF003105	RAP2.12U 5'-GCTTTGGTTCAGAACTCGAACATC-3' RAP2.12L 5'-AGGTTGATAAACGAACGATGCG-3'

Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, e.g., Carruthers *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams *et al.*, *J. Am. Chem. Soc.* 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the

complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Alternatively, primers that specifically hybridize to highly conserved regions in AP2 domains can be used to amplify sequences from widely divergent plant species such as *Arabidopsis*, canola, soybean, tobacco, and snapdragon. Examples of such primers are as follows:

Primer RISZU 1: 5'-GGAYTGTGGGAAACAAGTTTA-3'

Primer RISZU 2: 5'-TGCAAAGTRACACCTCTATACTT-3'

Y = pyrimidine (T or C)

R = purine (A or G).

Standard nucleic acid hybridization techniques using the conditions disclosed above can then be used to identify full length cDNA or genomic clones.

In addition, the following DNA primers, RISZU 3 and RISZU 4, can be used in an inverse PCR reaction to specifically amplify flanking AP2 gene sequences from widely divergent plant species. These primers are as follows:

Primer RISZU 3: 5'-GCATGWGCAGTGTCAAATCCA-3'

Primer RISZU 4: 5'-GAGGAAGTTCVAAGTATAGA-3'

W = A or T

V = G, A, or C

These primers have been used in standard PCR conditions to amplify ADC gene sequences from canola (SEQ ID NO:1) and soybean (SEQ ID NO:2).

### Suppression of ADC activity or gene expression

One of skill will recognize that a number of methods can be used to inactivate or suppress ADC activity or gene expression. The control of the expression can be achieved by introducing mutations into the gene or using recombinant DNA techniques. These techniques are generally well known to one of skill and are discussed briefly below.

Methods for introducing genetic mutations into plant genes are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are

not limited to, the following: diethyl sulfate, ethylene imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as, for example, X-rays or gamma rays can be used. Desired mutants are selected by assaying for increased seed mass, oil content and other properties.

5           Gene expression can be inactivated using recombinant DNA techniques by transforming plant cells with constructs comprising transposons or T-DNA sequences. *ADC* mutants prepared by these methods are identified according to standard techniques. For instance, mutants can be detected by PCR or by detecting the presence or absence of *ADC* mRNA, *e.g.*, by Northern blots. Mutants can also be selected by assaying for  
10           increased seed mass, oil content and other properties.

          The isolated sequences prepared as described herein, can also be used in a number of techniques to suppress endogenous *ADC* gene expression. A particularly useful genes for this purpose are the *AP2* gene described in Jofuku *et al.*, *supra*, and *RAP2* genes described below. For instance, sequences flanking the *AP2* domains of the genes  
15           disclosed here to specifically target individual genes.

          A number of methods can be used to inhibit gene expression in plants. For instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The construct is then transformed into  
20           plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, see, *e.g.*, Sheehy *et al.*, *Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt *et al.*, U.S. Patent No. 4,801,340.

          The nucleic acid segment to be introduced generally will be substantially  
25           identical to at least a portion of the endogenous *ADC* gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other genes within a family of genes exhibiting homology or substantial homology to the target gene.

          For antisense suppression, the introduced sequence also need not be full  
30           length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and

homology of non-coding segments may be equally effective. Normally, a sequence of between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of about 500 to about 1700 nucleotides is especially preferred.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of *ADC* genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs.

A number of classes of ribozymes have been identified. One class of ribozymes is derived from a number of small circular RNAs which are capable of self-cleavage and replication in plants. The RNAs replicate either alone (viroid RNAs) or with a helper virus (satellite RNAs). Examples include RNAs from avocado sunblotch viroid and the satellite RNAs from tobacco ringspot virus, lucerne transient streak virus, velvet tobacco mottle virus, solanum nodiflorum mottle virus and subterranean clover mottle virus. The design and use of target RNA-specific ribozymes is described in Haseloff et al. *Nature*, 334:585-591 (1988).

Another method of suppression is sense cosuppression. Introduction of nucleic acid configured in the sense orientation has been recently shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli et al., *The Plant Cell* 2:279-289 (1990), and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184.

The suppressive effect may occur where the introduced sequence contains no coding sequence *per se*, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous

sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

5 For sense suppression, the introduced sequence, needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid concurrent production of some plants which are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence  
10 need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used.

Some ADC proteins (e.g., AP2) are believed to form multimers *in vivo*. As a result, an alternative method for inhibiting ADC function is through use of dominant  
15 negative mutants. This approach involves transformation of plants with constructs encoding mutant ADC polypeptides that form defective multimers with endogenous wild-type ADC proteins and thereby inactivate the protein. The mutant polypeptide may vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a  
20 number of combinations to produce the final modified protein chain. Use of dominant negative mutants to inactivate AG is described in Mizukami *et al. Plant Cell* 8:831-845 (1996).

#### Use of nucleic acids of the invention to enhance ADC gene expression

25 Isolated sequences prepared as described herein can also be used to introduce expression of a particular ADC nucleic acid to enhance or increase endogenous gene expression. Enhanced expression will generally lead to smaller seeds or seedless fruit. Where overexpression of a gene is desired, the desired gene from a different species may be used to decrease potential sense suppression effects.

30 One of skill will recognize that the polypeptides encoded by the genes of the invention, like other proteins, have different domains which perform different functions. Thus, the gene sequences need not be full length, so long as the desired functional domain

of the protein is expressed. The distinguishing features of *ADC* polypeptides, including the AP2 domain, are discussed in detail below.

Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described in detail, below. For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

#### Preparation of recombinant vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. See, for example, Weising et al. *Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

For example, for overexpression, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill. Such genes include for example, the *AP2* gene, *ACT11* from *Arabidopsis* (Huang et al. *Plant Mol. Biol.* 33:125-139 (1996)), *Cat3* from *Arabidopsis* (GenBank No. U43147, Zhong et al., *Mol. Gen. Genet.* 251:196-203 (1996)), the gene encoding stearyl-acyl carrier protein desaturase from *Brassica napus* (Genbank No. X74782, Solocombe et al. *Plant Physiol.* 104:1167-1176 (1994)), *GPcI* from maize (GenBank No.

X15596, Martinez *et al. J. Mol. Biol.* 208:551-565 (1989)), and *Gpc2* from maize (GenBank No. U45855, Manjunath *et al., Plant Mol. Biol.* 33:97-112 (1997)).

Alternatively, the plant promoter may direct expression of the *ADC* nucleic acid in a specific tissue or may be otherwise under more precise environmental or developmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions, elevated temperature, or the presence of light. Such promoters are referred to here as "inducible" or "tissue-specific" promoters. One of skill will recognize that a tissue-specific promoter may drive expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a tissue-specific promoter is one that drives expression preferentially in the target tissue, but may also lead to some expression in other tissues as well.

Examples of promoters under developmental control include promoters that initiate transcription only (or primarily only) in certain tissues, such as fruit, seeds, or flowers. Promoters that direct expression of nucleic acids in ovules, flowers or seeds are particularly useful in the present invention. As used herein a seed-specific promoter is one which directs expression in seed tissues, such promoters may be, for example, ovule-specific, embryo-specific, endosperm-specific, integument-specific, seed coat-specific, or some combination thereof. Examples include a promoter from the ovule-specific *BEL1* gene described in Reiser *et al. Cell* 83:735-742 (1995) (GenBank No. U39944). Other suitable seed specific promoters are derived from the following genes: *MAC1* from maize (Sheridan *et al. Genetics* 142:1009-1020 (1996), *Cat3* from maize (GenBank No. L05934, Abler *et al. Plant Mol. Biol.* 22:10131-1038 (1993), the gene encoding oleosin 18kD from maize (GenBank No. J05212, Lee *et al. Plant Mol. Biol.* 26:1981-1987 (1994)), *viviparous-1* from *Arabidopsis* (Genbank No. U93215), the gene encoding oleosin from *Arabidopsis* (Genbank No. Z17657), *Atmyc1* from *Arabidopsis* (Urao *et al. Plant Mol. Biol.* 32:571-576 (1996), the 2s seed storage protein gene family from *Arabidopsis* (Conceicao *et al. Plant* 5:493-505 (1994)) the gene encoding oleosin 20kD from *Brassica napus* (GenBank No. M63985), *napA* from *Brassica napus* (GenBank No. J02798, Josefsson *et al. JBL* 26:12196-1301 (1987), the napin gene family from *Brassica napus* (Sjodahl *et al. Planta* 197:264-271 (1995), the gene encoding the 2S storage protein from *Brassica napus* (Dasgupta *et al. Gene* 133:301-302 (1993)), the genes encoding oleosin A (Genbank No. U09118) and oleosin B (Genbank No. U09119) from soybean and the gene

encoding low molecular weight sulphur rich protein from soybean (Choi *et al. Mol Gen, Genet.* 246:266-268 (1995)).

If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

The vector comprising the sequences (*e.g.*, promoters or coding regions) from genes of the invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

#### Production of transgenic plants

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski *et al. Embo J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm *et al. Proc. Natl. Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein *et al. Nature* 327:70-73 (1987).

Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria. *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific



literature. See, for example Horsch et al. *Science* 233:496-498 (1984), and Fraley et al. *Proc. Natl. Acad. Sci. USA* 80:4803 (1983).

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype such as increased seed mass. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans et al., *Protoplasts Isolation and Culture, Handbook of Plant Cell Culture*, pp. 124-176, MacMillan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants, Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee et al. *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

The nucleic acids of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from the genera *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panicum*, *Pannasetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Solanum*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna*, and *Zea*.

Increasing seed size, protein, amino acid, and oils content is particularly desirable in crop plants in which seed are used directly for animal or human consumption or for industrial purposes. Examples include soybean, canola, and grains such as rice, wheat, corn, rye, and the like. Decreasing seed size, or producing seedless varieties, is particularly important in plants grown for their fruit and in which large seeds may be undesirable. Examples include cucumbers, tomatoes, melons, and cherries.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into

other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

Since transgenic expression of the nucleic acids of the invention leads to phenotypic changes in seeds and fruit, plants comprising the expression cassettes discussed above must be sexually crossed with a second plant to obtain the final product. The seed of the invention can be derived from a cross between two transgenic plants of the invention, or a cross between a plant of the invention and another plant. The desired effects (*e.g.*, increased seed mass) are generally enhanced when both parental plants contain expression cassettes of the invention.

Seed obtained from plants of the present invention can be analyzed according to well known procedures to identify seed with the desired trait. Increased or decreased size can be determined by weighing seeds or by visual inspection. Protein content is conveniently measured by the method of Bradford *et al. Anal. Bioch.* 72:248 (1976). Oil content is determined using standard procedures such as gas chromatography. These procedures can also be used to determine whether the types of fatty acids and other lipids are altered in the plants of the invention.

Using these procedures one of skill can identify the seed of the invention by the presence of the expression cassettes of the invention and increased seed mass. Usually, the seed mass will be at least about 10%, often about 20% greater than the average seed mass of plants of the same variety that lack the expression cassette. The mass can be about 50% greater and preferably at least about 75% to about 100% greater. Increases in other properties *e.g.*, protein and oil will usually be proportional to the increases in mass. Thus, in some embodiments protein or oil content can increase by about 10%, 20%, 50%, 75% or 100%, or in approximate proportion to the increase in mass.

Alternatively, seed of the invention in which *AP2* expression is enhanced will have the expression cassettes of the invention and decreased seed mass. Seed mass will be at least about 20% less than the average seed mass of plants of the same variety that lack the expression cassette. Often the mass will be about 50% less and preferably at least about 75% less or the seed will be absent. As above, decreases in other properties *e.g.*, protein and oil will be proportional to the decreases in mass.

The following Examples are offered by way of illustration, not limitation.

## Example 1

*AP2* Gene Isolation

The isolation and characterization of an *AP2* gene from *Arabidopsis* is described in detail in Jofuku *et al.*, *supra*. Briefly, T-DNA from *Agrobacterium* was used as an insertional mutagen to identify and isolate genes controlling flower formation in *Arabidopsis*. One transformed line, designated T10, segregated 3 to 1 for a flower mutant that phenotypically resembled many allelic forms of the floral homeotic mutant *ap2*. T10 was tested and it was confirmed genetically that T10 and *ap2* are allelic. The mutant was designated as *ap2-10*.

It was determined that *ap2-10* was the product of a T-DNA insertion mutation by genetic linkage analysis using the T-DNA-encoded neomycin phosphotransferase II (NPTII) gene as a genetic marker. An overlapping set of T-DNA-containing recombinant phage was selected from an *ap2-10* genome library and the plant DNA sequences flanking the T-DNA insertion element were used as hybridization probes to isolate phage containing the corresponding region from a wild-type *Arabidopsis* genome library. The site of T-DNA insertion in *ap2-10* was mapped to a 7.2-kb *EcoRI* fragment centrally located within the *AP2* gene region.

Five *Arabidopsis* flower cDNA clones corresponding to sequences within the 7.2-kb *AP2* gene region were isolated. All five cloned cDNAs were confirmed to represent *AP2* gene transcripts using an antisense gene strategy to induce *ap2* mutant flowers in wild-type plants.

To determine *AP2* gene structure, the nucleotide sequences of the cDNA inserts were compared to that of the 7.2-kb *AP2* genomic fragment. These results showed that the *AP2* gene is 2.5 kb in length and contains 10 exons and 9 introns that range from 85 to 110 bp in length. The *AP2* gene encodes a theoretical polypeptide of 432 amino acids with a predicted molecular mass of 48 kD. The *AP2* nucleotide and predicted protein sequences were compared with a merged, nonredundant data base. It was found that *AP2* had no significant global similarity to any known regulatory protein.

Sequence analysis, however, did reveal the presence of several sequence features that may be important for *AP2* protein structure or function. First, *AP2* contains a 37-amino acid serine-rich acidic domain (amino acids 14 to 50) that is analogous to regions that function as activation domains in a number of RNA polymerase II

transcription factors. Second, *AP2* has a highly basic 10-amino acid domain (amino acids 119 to 128) that includes a putative nuclear localization sequence KKSR suggesting that *AP2* may function in the nucleus. Finally, that the central core of the *AP2* polypeptide (amino acids 129 to 288) contains two copies of a 68-amino acid direct repeat that is referred to here as the *AP2* domain. The two copies of this repeat, designated *AP2-R1* and *AP2-R2*, share 53% amino acid identity and 69% amino acid homology. Figure 1A shows that each *AP2* repeat contains an 18-amino acid conserved core region that shares 83% amino acid homology. Figure 1B shows that both copies of this core region are theoretically capable of forming amphipathic  $\alpha$ -helical structures that may participate in protein-protein interactions.

## Example 2

### Preparation of *AP2* Constructs

Gene constructs were made comprising the *AP2* gene coding region described above in a transcriptional fusion with the cauliflower mosaic virus 35S constitutive promoter in both the sense and antisense orientations. The original vector containing the 35S promoter pGSJ780A was obtained from Plant Genetic Systems (Gent, Belgium). The pGSJ780A vector was modified by inserting a ClaI-BamHI adaptor containing an EcoRI site in the unique BamHI site of pGSJ780A. The modified pGSJ780A DNA was linearized with EcoRI and the *AP2* gene coding region inserted as a 1.68 kb EcoRI fragment in both sense and antisense orientations with respect to the 35S promoter (see, Figures 2 and 3).

The resultant DNA was transformed into *E. coli* and spectinomycin resistant transformants were selected. Plasmid DNAs were isolated from individual transformants and the orientation of the insert DNAs relative to the 35S promoter were confirmed by DNA sequencing. Bacterial cells containing the 35S/*AP2* sense (designated pPW12.4 and pPW9) and 35S/*AP2* antisense (designated pPW14.4 and pPW15) constructs were conjugated to *Agrobacterium tumefaciens* and rifampicin, spectinomycin resistant transformants were selected for use in *Agrobacterium*-mediated plant transformation experiments.

The 35S/*AP2* sense and 35S/*AP2* antisense constructs were introduced into wild-type *Arabidopsis* and tobacco plants according to standard techniques. Stable

transgenic plant lines were selected using the plant selectable marker *NPTII* (which confers resistance to the antibiotic kanamycin) present on the modified Ti plasmid vector pGSJ780A.

### Example 3

#### Modification of Seed using *AP2* Sequences

This example shows that *ap2* mutant plants and transgenic plants containing the 35S/*AP2* antisense construct produced seed with increased mass and total protein content. By contrast, transgenic plants containing the 35S/*AP2* sense construct produced seed with decreased mass and protein content. Together these results indicate that seed mass and seed contents in transgenic plants can be modified by genetically altering *AP2* activity.

Seed from 30 lines were analyzed for altered seed size and seed protein content including the Arabidopsis *ap2* mutants *ap2-1*, *ap2-3*, *ap2-4*, *ap2-5*, *ap2-6*, *ap2-9* and *ap2-10* and transgenic Arabidopsis and transgenic tobacco containing the CaMV 35S/*AP2* antisense gene construct, the CaMV 35S/*AP2* sense gene construct, or the pGSJ780A vector as described above. The *ap2* mutants used in this study are described in Komaki *et al.*, *Development* 104, 195-203 (1988), Kunst *et al.*, *Plant Cell* 1, 1195-1208 (1989), Bowman *et al.*, *Development* 112, 1-20 (1991), and Jofuku *et al.*, *supra*.

Due to the small size of Arabidopsis and tobacco seed, average seed mass was determined by weighing seed in batches of 100 for Arabidopsis and 50 seed for tobacco. The net change in seed mass due to changes in *AP2* gene activity was calculated by subtracting the average mass of wild-type seed from mutant seed mass.

Seed from three wild-type Arabidopsis ecotypes C24, Landsberg-*er*, and Columbia, and one wild-type tobacco SR1 were used as controls. Wild-type Arabidopsis seed display seasonal variations in seed mass which range from 1.6-2.3 mg per 100 seed as shown in Table I. Therefore transgenic Arabidopsis seed were compared to control seed that had been harvested at approximately the same time of season. This proved to be an important for comparing the effects of weak *ap2* mutations on seed mass.

Table I shows that all *ap2* mutant seed examined, *ap2-1*, *ap2-3*, *ap2-4*, *ap2-5*, *ap2-6*, *ap2-9*, and *ap2-10*, show a significant increase in average seed mass ranging from +27 to +104 percent compared to wild-type. The weak partial loss-of-function

mutants such as *ap2-1* and *ap2-3* show the smallest gain in average seed mass ranging from +27 percent to +40 percent of wild-type, respectively. By contrast, strong *ap2* mutants such as *ap2-6* and *ap2-10* show the largest gain in seed mass ranging from +69 percent to +104 percent of wild-type, respectively. Thus reducing *AP2* gene activity genetically consistently increases *Arabidopsis* seed mass.

*AP2* antisense and *AP2* sense cosuppression strategies described above were used to reduce *AP2* gene activity *in planta* to determine whether seed mass could be manipulated in transgenic wild-type plants. Twenty-nine independent lines of transgenic *Arabidopsis* containing the CaMV 35S/*AP2* antisense gene constructs pPW14.4 and pPW15 (Figure 2) were generated. Each transgenic line used in this study tested positive for kanamycin resistance and the presence of one or more copies of T-DNA.

Table I shows that seed from nine transgenic *Arabidopsis* *AP2* antisense lines show a significant increase in seed mass when compared to control seed ranging from +22 percent for line C24 15-542 to +89 percent for line C24 15-566. Both C24 and Landsberg-*er* ecotypes were used successfully. Increased seed mass was observed in F1, F2, and F3 generation seed.

Eight lines containing the 35S/*AP2* sense gene construct were generated which were phenotypically cosuppression mutants. As shown in Table I seed from two cosuppression lines examined showed larger seed that range from +26 percent to +86 percent. By contrast, plants transformed with the vector pGSJ780A showed a normal range of average seed mass ranging from -0.5 percent to +13 percent compared to wild-type seed (Table I). Together, these results demonstrate that *AP2* gene sequences can be used to produce a significant increase in *Arabidopsis* seed mass using both antisense and cosuppression strategies in a flowering plant.

Table I. Genetic control of Arabidopsis seed mass by AP2.

		Average seed mass in mg per 100 seed <sup>1,2</sup>	Percent change in seed mass compared to wild-type
ap2 mutant seed		2.1 (0.1)	+27%
5	1. ap2-1	2.2 (0.1)	+33%
		2.1 (0.2)	+31%
		2.8 (0.2)	+33%
	2. ap2-3	2.6 (0.1)	+27%
	3. ap2-4	3.5 (0.3)	+69%
10		3.5 (0.2)	+69%
	4. ap2-5	2.9 (0.1)	+39%
	5. ap2-6	3.5 (0.2)	+69%
	6. ap2-9	2.9 (0.1)	+40%
	7. ap2-10	3.7 (0.4)	+79%
15		3.9 (0.3)	+90%
		4.2 (0.5)	+104%
Seed produced by transgenic CaMV35S/AP2 antisense lines (from a Km resistant mother)			
20	1. C24 14.4E (F1-15) F2 sd	3.1	+35%
	C24 14.4E (F1-15) F3 sd	3.4 (0.3)	+47%
	2. C24 14.4S (F1-1)	2.8 (0.2)	+29%
	3. C24 14.4AA (F1-24)	2.9 (0.1)	+30%
	4. C24 14.4DD (F1-2)	2.8 (0.3)	+30%
25	5. C24 15-522	3.6 (0.1)	+76%
	6. C24 15-542 (F1-2)	2.6 (0.1)	+25%
	C24 15-542 (F1-7)	2.5 (0.2)	+22%
	7. C24 15-566	3.9 (0.1)	+89%
	8. LE 15-9992-3 (F1-1) F2 sd	2.4 (0.1)	+42%
30	9. LE 15-83192-3 (F1-3)	2.8 (0.0)	+33%
	LE 15-83192-3 (F1-17)	2.7 (0.0)	+28%
Seed produced by transgenic CaMV35S/AP2 cosuppression lines (from a Km resistant mother)			
30	1. C24 9-5 (F1-5)	3.8 (0.0)	+86%
	2. LE 9-83192-2 (F1-19)	2.7 (0.2)	+26%
	LE 9-83192-2 (F1-24)	2.7 (0.1)	+26%

Average seed mass in		Percent change in seed mass compared to wild-type
mg per 100 seed <sup>1,2</sup>		
Seed produced by transgenic pGSJ780A vector only lines (from a Km resistant mother plant)		
1. C24 3-107 (F1-1)	2.2 (0.1)	+ 9%
2. C24 3-109 (F1-1)	2.3 (0.0)	+ 13%
3. LE 3-83192-1 (F1-2)	2.3 (0.1)	+ 7%
4. LE 3-83192-3 (F1-2)	2.4 (0.1)	+ 11%
5. LE 3-9992-4 (F1-4)	2.4 (0.2)	+ 12%
LE 3-9992-4 (F1-6)	2.3 (0.0)	+ 9%
LE 3-9992-4 (F1-8)	2.1 (0.0)	-0.5%
6. LE 3-9992-9 (F1-3)	2.3 (0.1)	+ 7%
Seed produced by wild-type Arabidopsis plants		
1. C24	2.0 (0.1)	
	2.3 (0.1)	
	2.2	
2. Landsberg- <i>er</i>	1.6 (0.1)	
	2.1 (0.1)	
	2.1	
	2.3 (0.1)	
3. Columbia	1.8 (0.1)	
	2.1 (0.1)	

<sup>1</sup> Standard deviation values are given in parentheses.

<sup>2</sup> Wild-type seed values used for this comparison were chosen by ecotype and harvest date.

Arabidopsis *AP2* gene sequences were also used to negatively control seed mass in tobacco, a heterologous plant species. Table II shows that in five transgenic tobacco lines the CaMV 35S/*AP2* overexpression gene construct was effective in reducing transgenic seed mass from -27 percent to -38 percent compared to wild-type seed. These results demonstrate the evolutionary conservation of *AP2* gene function at the protein level for controlling seed mass in a heterologous system.



Table II. Genetic control of tobacco seed mass using Arabidopsis AP2.

		Average seed mass in mg per 5 seed <sup>1</sup>	Percent change in seed mass compared to wild-type
Seed produced by transgenic CaMV 35S/AP2 sense gene lines (from a Km resistant mother)			
5	1. SR1 9-110 To	3.1 (0.0)	-27%
	SR1 9-110 (F1-5)	3.0 (0.2)	-29%
		2.8 (0.3)	-34%
10	2. SR1 9-202 (F1-G)	3.1 (0.2)	-27%
	SR1 9-202 (F1-I)	3.2 (0.1)	-24%
	3. SR1 9-103 (F1-2)	3.9 (0.0)	-8%
	4. SR1 9-413-1	2.8 (0.0)	-34%
		3.0 (0.2)	-29%
	5. SR1 9-418-1 To	3.5 (0.1)	-18%
Seed produced by transgenic CaMV 35S/AP2 antisense gene lines (from a Km resistant mother)			
15	1. SR1 15-111	5.1 (0.4)	+20%
	SR1 15-111 (F1)	5.0 (0.4)	+19%
	2. SR1 15-116 To	4.1 (0.4)	-3%
20	SR1 15-116 (F1-2)	4.0 (0.1)	-5%
	SR1 15-116 (F1-1)	4.5 (0.1)	+5%
	3. SR1 15-407 (F1)	4.8 (0.5)	+10%
		4.7 (0.3)	+10%
	4. SR1 15-102 (F1)	4.5 (0.2)	+6%
25	5. SR1 15-413 (F1-3)	4.2 (0.0)	+0%
	6. SR1 15-410 (F1-2)	4.4 (0.0)	+4%
	7. SR1 15-210 (F1-4)	3.6 (0.1)	-15%
Seed produced by pGSJ780A vector only lines (from a Km resistant mother)			
30	1. SR1 3-402 (F1)	5.0 (0.1)	+17%
	2. SR1 3-401 (F1)	4.6 (0.1)	+8%
	3. SR1 3-405 (F1)	4.4 (0.1)	+4%
Seed from wild-type tobacco			
	1. SR1	4.2 (0.3)	
		4.0 (0.1)	

<sup>1</sup> Standard deviation values are given in parentheses.

### Use of AP2 gene constructs to control seed protein content

Total seed protein was extracted and quantitated from seed produced by wild-type, ap2 mutant, transgenic AP2 antisense, and transgenic AP2 sense cosuppression plants according to Naito *et al. Plant Mol Biol.* 11, 109-123 (1988). Seed protein was extracted in triplicate from batches of 100 dried seed for Arabidopsis or 50 dried seed for tobacco. Total protein yield was determined by the Bradford dye-binding procedure as

described by Bradford, *Anal. Biochem.* 72:248 (1976). The results of this analysis are shown in Table III.

*ap2* mutant total seed protein content increased by 20 percent to 78 percent compared to wild-type control seed. Total seed protein from transgenic *AP2* antisense plants increased by +31 percent to +97 percent compared to wild-type controls. Transgenic *AP2* cosuppression seed showed a +13 and +17 percent increase over wild-type. Together, the transgenic antisense and cosuppression mutant seed consistently yielded more protein per seed than did the wild-type controls or transgenic plants containing the pGSJ780A vector only (Table III).

Table III. Genetic control of total seed protein content in *Arabidopsis* using *AP2*.

	Total seed protein in $\mu\text{g}$ per 100 seed <sup>1</sup>	Percent change in protein content compared to wild-type
<i>ap2</i> mutant seed		
1. <i>ap2-1</i>	652 (17)	+20% relative to WT seed
	615 (30)	+11%
2. <i>ap2-3</i>	705 (47)	+27%
3. <i>ap2-4</i>	729 (107)	+33%
4. <i>ap2-5</i>	617 (24)	+13%
5. <i>ap2-6</i>	836 (14)	+52%
6. <i>ap2-9</i>	798 (11)	+46%
7. <i>ap2-10</i>	836 (15)	+78%
Transgenic CaMV 35S/ <i>AP2</i> antisense seed mass (from Km resistant mother)		
1. C24 14.4E (F1-1) F3 sd	615 (60)	+31%
2. C24 15-522 (F1-1)	790 (23)	+68%
3. C24 15-566	925 (173)	+97%
Transgenic CaMV 35S/ <i>AP2</i> sense cosuppression seed mass (from Km Resistant mother plant)		
1. LE 9-83192-2 (F1-19)	616	+13%
LE 9-83192-2 (F1-24)	637	+17%
Wild-type seed		
1. C24	469 (19)	
2. LE	545 (22)	
	555	
3. Col	548 (42)	

<sup>1</sup> Standard deviation values are given in parentheses.

Transgenic tobacco containing the 35S/*AP2* sense gene construct show that *AP2* overexpression can decrease seed protein content by 27 to 45 percent compared to

wild-type seed. Together, the transgenic Arabidopsis and tobacco results demonstrate that seed mass and seed protein production can be controlled by regulating AP2 gene activity.

Table IV. Negative control of transgenic tobacco seed protein content by Arabidopsis AP2 gene expression.<sup>1</sup>

	Ave. protein per 50 seed	Percent change in protein content compared to wild-type
Seed produced by transgenic CAMV 35S/AP2 sense gene plant		
1. SR1 9-110	242 (11)	-45%
2. SR1 9-202 (F1-G)	271 (11)	-38%
3. SR1 9-413	362 (8)	-18%
4. SR1 9-418-1	319 (16)	-27%
Wild-type Control SR1 (wild-type)	440 (8) (JO)	NA

<sup>1</sup> Standard deviation values are given in parentheses.

#### Analysis of transgenic seed proteins by gel electrophoresis

Arabidopsis seed produce two major classes of seed storage proteins, the 12S cruciferins and 2S napins which are structurally related to the major storage proteins found in the *Brassicaceae* and in the *Leguminosae*. The composition of seed proteins in wild-type, *ap2* mutant, and transgenic Arabidopsis seed were compared by SDS polyacrylamide gel electrophoresis as described by Naito *et al.*, *Plant Mol. Biol.* 11, 109-123 (1988). Total seed proteins were extracted as described above. 50 µg aliquots were separated by gel electrophoresis and stained using Coomassie brilliant blue. These results showed that the spectrum of proteins in wild-type and *ap2* mutant seed are qualitatively indistinguishable. There is no detectable difference in the representation of the 12S or 2S storage proteins between the wild-type and *ap2* mutant seed extracts. This shows that reducing AP2 gene activity genetically does not alter the profile of storage proteins synthesized during seed maturation. The spectrum of seed proteins produced in transgenic AP2 antisense and AP2 sense cosuppression seed are also indistinguishable from wild-type. In particular, there is no detectable difference in the representation of the 12S cruciferin or 2S napin storage proteins in the larger seed.

Finally, the transgenic tobacco plants containing the 35S/AP2 overexpression gene construct produced significantly smaller seed. Despite the decrease in

seed mass in transgenic tobacco there was no detectable difference in storage protein profiles between seed from 35S/AP2 transformants and wild-type SR1.

#### Example 4

Isolation of other members of the AP2 gene family from *Arabidopsis*

This example describes isolation of a number of AP2 nucleic acids from *Arabidopsis*. The nucleic acids are referred to here as RAP2 (related to AP2) were identified using primers specific to nucleic acid sequences from the AP2 domain described above.

#### MATERIALS AND METHODS

**Plant Material.** *Arabidopsis thaliana* ecotype Landsberg *erecta* (L-er) and C24 were used as wild type. Plants were grown at 22°C under a 16-hr light/8-hr dark photoperiod in a 1:1:1 mixture containing vermiculite/perlite/peat moss. Plants were watered with a one-fourth strength Peter's solution (Grace-Sierra, Milpitas, CA). Root tissue was harvested from plants grown hydroponically in sterile flasks containing 1x Murashige and Skoog plant salts (GIBCO), 1 mg/liter thiamine, 0.5 mg/liter pyridoxine, 0.5 mg/liter nicotinic acid, 0.5 g/liter 2-(N-morpholino)ethanesulfonic acid (MES), and 3% sucrose, with moderate shaking and 70  $\mu\text{mol} \cdot \text{m}^{-2} \cdot \text{sec}^{-1}$  of light.

**Analysis of cloned *Arabidopsis* cDNAs.** *Arabidopsis* expressed sequence tagged (EST) cDNA clones representing RAP2.1 and RAP2.9 were generated as described by Cooke *et al.* (Cooke, R., *et al.*, 1996, *Plant J.* 9, 101-124). EST cDNA clones representing RAP2.2 and RAP2.8 were generated as described by Höfte *et al.* (Höfte, H., *et al.*, 1993, *Plant J.* 4, 1051-1061). EST cDNA clones representing all other RAP2 genes were generated by Newman *et al.* (Newman, T., *et al.*, 1994, *Plant Physiol.* 106, 1241-1255) and provided by the *Arabidopsis* Biological Resource Center (Ohio State University). Plasmid DNAs were isolated and purified by anion exchange chromatography (Qiagen, Chatsworth, CA). DNA sequences were generated using fluorescence dye-based nucleotide terminators and analyzed as specified by the manufacturer (Applied Biosystems).

**Nucleotide and Amino Acid Sequence Comparisons.** The TBLASTN program (Altschul, S. F., *et al.*, 1990, *J. Mol. Biol.* 215, 403-410) and default parameter

settings were used to search the *Arabidopsis* EST database (AAtdB 4-7) for genes that encode AP2 domain-containing proteins. Amino acid sequence alignments were generated using the CLUSTAL W multiple sequence alignment program (Thompson, J. D., *et al.*, 1994, *Nucleic Acids Res.* 22, 4673-4680). Secondary structure predictions were based on the principles and software programs described by Rost (Rost, B., 1996, *Methods Enzymol.* 266, 525-539) and Rost and Sander (Rost, B., *et al.*, 1993, *J. Mol. Biol.* 232, 589-599; Rost, B., *et al.*, 1994, *Proteins* 19, 55-77).

**RAP2 Gene-Specific Probes.** RAP2 gene-specific fragments were generated by PCR using gene-specific primers and individual RAP2 plasmid DNAs as a template as specified by Perkin-Elmer (Roche Molecular Systems, Branchburg, NJ). The following primers were used to generate fragments representing each RAP2 gene:

RAP2.1, 5'-AAGAGGACCATCTCTCAG-3',

5'-AACACTCGCTAGCTTCTC-3';

RAP2.2, 5'-TGGTTCAGCAGCCAACAC-3', 5'-CAATGCATAGAGCTTGAGG-3';

RAP2.3, 5'-TCATCGCCACGATCAACC-3', 5'-AGCAGTCCAATGCGACGG-3';

RAP2.4, 5'-ACGGATTTACATCGGAG-3', 5'-CTAAGCTAGAATCGAATCC-3';

RAP2.7, 5'-CGATGGAGACGAAGACTC-3, 5'-GTCGGAACCGGAGTTACC-3';

RAP2.8, 5'-TCACTCAAAGGCCGAGATC-3', 5'-TAACAACATCACC GGCTCG-3';

RAP2.9, 5'-GTGAAGGCTTAGGAGGAG-3', 5'-TGCCTCATATGAGTCAGAG-3'.

PCR-synthesized DNA fragments were gel purified and radioactively labeled using random oligonucleotides (Amersham) for use as probes in gene mapping and RNA gel blot experiments.

**Gene Mapping Experiments.** RAP2 genes were placed on the *Arabidopsis* genetic map by either restriction fragment length polymorphism segregation analysis using recombinant inbred lines as described by Reiter *et al.* (Reiter, R. S., *et al.*, 1992, *Proc. Natl. Acad. Sci. USA* 89, 1477-1481) or by matrix-based analysis of pooled DNAs from the *Arabidopsis* yUP or CIC yeast artificial chromosome (YAC) genomic libraries (Ecker, J. R., 1990, *Methods* 1, 186-194; Creusot, F., *et al.*, 1995, *Plant J.* 8, 763-770) using the PCR (Green, E. D., *et al.*, 1990, *Proc. Natl. Acad. Sci. USA* 87, 1213-1217; Kwiatkowski, T. J., *et al.*, 1990, *Nucleic Acids Res.* 18, 7191-7192). Matrix based mapping results were confirmed by PCR using DNA from individual YAC clones.

**mRNA Isolation.** Polysomal poly(A) mRNAs from *Arabidopsis* flower, rosette leaf, inflorescence stem internode, and hydroponically-grown roots were isolated according to Cox and Goldberg (Cox, K. H., *et al.*, 1988, in *Plant Molecular Biology. A Practical Approach*, ed. Shaw, C. H. (IRL, Oxford), pp. 1-35).

**RNA Gel Blot Studies.** RNA gel blot hybridizations were carried out as specified by the manufacturer (Amersham). mRNA sizes were estimated relative to known RNA standards (BRL). AP2 transcripts were detected using a labeled DNA fragment representing nucleotides 1-1371 of the AP2 cDNA plasmid clone pAP2cl (Jofuku, K. D., *et al.*, 1994, *Plant Cell* 6, 1211-1225).

## RESULTS

**The AP2 Domain Defines a Large Family of Plant Proteins.** Using the AP2 domain as a sequence probe 34 cDNA clones were identified that encode putative RAP2 proteins in the *Arabidopsis* EST database (*Materials and Methods*). Several of these partial sequences have been reported previously (Ohme-Takagi, *et al.*, 1995, *Plant Cell* 7, 173-182; Elliot, R. C., *et al.*, 1996, *Plant Cell* 8, 155-168; Klucher, K. M., *et al.*, 1996, *Plant Cell* 8, 137-153; Wilson, K., *et al.*, 1996, *Plant Cell* 8, 659-671; Ecker, J. R., 1995, *Science* 268, 667-675; Weigel, D., 1995, *Plant Cell* 7, 388-389). Based on nucleotide sequence comparison, it was inferred that approximately half of the 34 RAP2 cDNA sequences were likely to represent redundant clones. Therefore, a complete DNA sequence for 17 putative RAP2 cDNA clones that appeared to represent unique genes and which contained the largest cDNA inserts was selected and generated. It was determined from the predicted amino acid sequences of these clones that the *Arabidopsis* RAP2 ESTs represent a minimum of 12 genes that are designated *RAP2.1-RAP2.12*. As shown in Table V, preliminary gene mapping experiments using restriction fragment length polymorphism analysis and PCR-based screening of the *Arabidopsis* yUP and CIC yeast artificial chromosome libraries (*Materials and Methods*) revealed that at least 7 members of the RAP2 gene family are distributed over 4 different chromosomes. In addition, several family members are tightly linked in the genome. For example, *RAP2.10* is only 10 kb away from *AP2*, which is also closely linked to *ANT* on chromosome 4 (Elliot, R.

C., *et al.*, 1996, *Plant Cell* 8, 155-168; Klucher, K M., *et al.*, 1996, *Plant Cell* 8, 137-153).

Sequence analysis also revealed that the proteins encoded by the RAP2 genes are all characterized by the presence of at least one AP2 domain. Fig. 4 shows a sequence comparison of 21 AP2 domains from 19 different polypeptides including RAP2.1-RAP2.12, AP2, ANT, TINY, and the tobacco EREBPs. From this comparison, it was determined that there are 2 conserved sequence blocks within each AP2 domain. The first block, referred to as the YRG element, consists of 19-22 amino acids, is highly basic and contains the conserved YRG amino acid motif (Fig. 4 A and B). The second block, referred to as the RAYD element, is 42-43 amino acids in length and contains a highly conserved 18-amino acid core region that is predicted to form an amphipathic  $\alpha$ -helix in the AP2 domains of AP2, ANT, TINY, and the EREBPs. In addition, there are several invariant amino acid residues within the YRG and RAYD elements that may also play a critical role in the structure or function of these proteins. For example, the glycine residue at position 40 within the RAYD element is invariant in all AP2 domain containing proteins (Fig. 4 A and B) and has been shown to be important for AP2 function (Jofuku, K. D., *et al.*, 1994, *Plant Cell* 6, 1211-1225).

Table V. *Arabidopsis* RAP2 genes

Gene	RAP2 gene containing YAC clones*	Chromosome map position†
<i>AINTEGUMENTA</i>	ND	4-73
<i>TINY</i>	ND	5-32 to 5-45
<i>RAP2.1</i>	yUP18H2, CIC11D10	ND‡
<i>RAP2.2</i>	yUP6C1 yUP12G6, yUP24B8, yUP23E11, CIC4H5,	38§
<i>RAP2.3</i>	CIC12C2	3-21
<i>RAP2.4</i>	CIC7D2, CIC10C4	ND‡
<i>RAP2.7</i>	yUP10E1	ND‡
<i>RAP2.8</i>	CIC10G7	1-94 to 1-103§
<i>RAP2.9</i>	CIC9E12	1-117§
<i>RAP2.10</i> <sup>1</sup>	ND	4-73

\* YAC clones were determined to contain the specified RAP2 gene by PCR-based DNA synthesis using gene-specific primers (Green, E. D., *et al.*, 1990, *Proc. Natl. Acad. Sci. USA* 87, 1213-1217; Kwiatkowski, T. J., *et al.*, 1990, *Nucleic Acids Res.* 18, 7191-7192).

† Chromosome map positions are given with reference to the *Arabidopsis* unified genetic map (AAIDB 4-7).

‡ YAC-based map position is ambiguous.

§ Preliminary map position is based on a single contact with the physical map.

GenBank accession numbers for complete EST sequences for RAP2 and other genes are as follows: *AINTEGUMENTA* (U40256/U41339); *TINY*, (X94598), *RAP2.1* (AF003094), *RAP2.2* (AF003095), *RAP2.3* (AF003096), *RAP2.4* (AF003097), *RAP2.5* (AF003098), *RAP2.6* (AF003099), *RAP2.7* (AF003100), *RAP2.8* (AF003101), *RAP2.9* (AF003102), *RAP2.10* (AF003103), *RAP2.11* (AF003104), and *RAP2.12* (AF003105). All RAP2 cDNA clones were originally reported with partial sequences and given GenBank accession numbers as shown in parentheses following each gene name: *RAP2.1* (Z27045), *RAP2.2* (Z26440), *RAP2.3* (T04320 and T13104), *RAP2.4* (T13774), *RAP2.5* (T45365), *RAP2.6* (T45770), *RAP2.7* (T20443), *RAP2.8* (Z33865), *RAP2.9* (Z37270), *RAP2.10* (T76017), *RAP2.11* (T42962), and *RAP2.12* (T42544). Due to the preliminary



nature of the EST sequence data, the predicted amino acid sequences for EST Z27045, T04320, T13774, and T42544 contained several errors and were incorrectly reported (Ohme-Takagi, *et al.*, 1995, *Plant Cell* 7, 173-182; Klucher, K M., *et al.*, 1996, *Plant Cell* 8, 137-153; Wilson, K., *et al.*, 1996, *Plant Cell* 8, 659-671; Ecker, J. R., 1995, *Science* 268, 667-675; Weigel, D., 1995, *Plant Cell* 7, 388-389). They are correctly given in the GenBank accession numbers noted above.

RAP2 cDNA sequence comparison also shows that there are at least two branches to the RAP2 gene family tree. The AP2-like and EREBP-like branches are distinguished by the number of AP2 domains contained within each polypeptide and by sequences within the conserved YRG element. The AP2-like branch of the RAP2 gene family is comprised of three genes *AP2*, *ANT*, and *RAP2.7*, each of which encodes a protein containing two AP2 domains (Fig. 4A). In addition, these proteins possess a conserved WEAR/WESH amino acid sequence motif located in the YRG element of both AP2 domain repeats (Fig. 4A). By contrast, genes belonging to the EREBP-like branch of the RAP2 gene family encode proteins with only one AP2 domain and include *RAP2.1-RAP2.6*, *RAP2.8-RAP2.12*, and *TINY* (Fig. 4B). Proteins in this class possess a conserved 7-amino acid sequence motif referred to as the WAAEIRD box (Fig. 4B) in place of the WEAR/WESH motif located in the YRG element (Fig. 4A). Based on these comparisons, separate AP2 domain consensus sequences for both classes of RAP2 proteins were generated (Fig. 4 A and B). These results suggest that the AP2 domain and specific sequence elements within the AP2 domain are important for RAP2 protein functions.

The AP2-like class of RAP2 proteins is also characterized by the presence of a highly conserved 25-26 amino acid linker region that lies between the two AP2 domain repeats (Klucher, K M., *et al.*, 1996, *Plant Cell* 8, 137-153). This region is 40% identical and 48% similar between AP2, ANT and RAP2.7 and is not found in proteins belonging to the EREBP-like branch of RAP2 proteins. Molecular analysis of the *ant-3* mutant allele showed that the invariant C-terminal glycine residue within this linker region is essential for ANT function *in vivo* (Klucher, K M., *et al.*, 1996, *Plant Cell* 8, 137-153), suggesting that the linker region may also play an important role in AP2 and RAP2.7 function.

**Sequences Within the RAYD Element are Predicted to Form Amphipathic  $\alpha$ -Helices.** As noted above, the 18-amino acid core region within the RAYD element of

the AP2 domain in AP2 is predicted to form an amphipathic  $\alpha$ -helix that may be important for AP2 structure or function. Secondary structure prediction analysis was used to determine whether this structure has been conserved in RAP2 proteins. As shown in Fig. 4, the core region represents the most highly conserved sequence block in the RAYD element of AP2 and the RAP2 proteins. Secondary structure analysis predicts that all RAP2 proteins contain sequences within the RAYD element that are predicted to form amphipathic  $\alpha$ -helices (Fig. 4 A and B). Fig. 4C shows that sequences in RAP2.7-R1 are predicted to form an amphipathic  $\alpha$ -helix that is 100% identical to that predicted for AP2-R1 and 63% similar to that predicted for ANT-R1. Sequences within the AP2 domain of EREBP-like RAP2 proteins are predicted to form similar  $\alpha$ -helical structures. Fig. 4D shows that the RAP2.2, RAP2.5, and RAP2.12  $\alpha$ -helices are 81, 100, and 81% similar to that predicted for EREBP-3, respectively. Together, these results strongly suggest that the predicted amphipathic  $\alpha$ -helix in the RAYD element is a conserved structural motif that is important for AP2 domain function in all RAP2 proteins.

**RAP2 Genes are Expressed in Floral and Vegetative Tissues.** Previous studies have shown that *AP2* and *ANT* are differentially expressed at the RNA level during plant development (Jofuku, K. D., *et al.*, 1994, *Plant Cell* 6, 1211-1225; Elliot, R. C., *et al.*, 1996, *Plant Cell* 8, 155-168; Klucher, K M., *et al.*, 1996, *Plant Cell* 8, 137-153). *AP2* is expressed at different levels in developing flowers, leaves, inflorescence stems, and roots. To determine where in plant development the EREBP-like class of RAP2 genes are expressed *RAP2.1*, *RAP2.2*, *RAP2.3*, and *RAP2.4* gene-specific probes were reacted with a mRNA gel blot containing flower, leaf, inflorescence stem, and root polysomal poly(A) mRNA. Results from these experiments showed that each RAP2 gene produces a uniquely sized mRNA transcript and displays a distinct pattern of gene expression in flowers, leaves, inflorescence stems, and roots. For example, the *RAP2.1* gene is expressed at low levels in wild-type flower, leaf, stem, and root. *RAP2.2* gene expression appears to be constitutive in that *RAP2.2* transcripts are detected at similar levels in wild-type flower, leaf, stem, and root. By contrast, the *RAP2.3* gene is expressed at a low level in wild-type flowers, at a slightly higher level in leaves, and is relatively highly expressed in both stems and roots. Finally, the *RAP2.4* gene is also expressed in wild-type flower, leaf, stem, and root and is most highly expressed in roots and leaves. These data indicate that individual members of the EREBP-like family of RAP2 genes are expressed at the mRNA

level in both floral and vegetative tissues and show quantitatively different patterns of gene regulation.

**RAP2 Gene Expression Patterns are Affected by *ap2*.** RAP2 gene expression was analyzed in *ap2-10* mutant plants by RNA gel blot analysis to determine whether AP2 is required for RAP2 gene expression. The expression of three RAP2 genes are differentially affected by the loss of AP2 function. For example, *RAP2.2* gene expression is not dramatically altered in mutant flowers, leaves, and roots compared to wild-type Landsberg *erecta* but is down-regulated in mutant stem. *RAP2.3* gene expression appears unchanged in mutant roots but is up-regulated in mutant flowers and leaves and down-regulated in mutant stems. By contrast, *RAP2.4* gene expression appears relatively unchanged in mutant stems and roots but is slightly up-regulated in mutant flowers and leaves. To control for possible secondary effects of ecotype on RAP2 gene expression, RAP2 gene expression levels in wild-type C24 and *ap2-10* mutant stems were compared. These results show that the differences in *RAP2.2* *RAP2.3*, and *RAP2.4* gene expression in C24 and *ap2-10* stem are similar to those observed between wild-type Landsberg *erecta* and *ap2-10* mutant stem. Together these results suggest that AP2 directly or indirectly regulates the expression of at least three RAP2 genes. More importantly, these results suggest that AP2 is controlling gene expression during both reproductive and vegetative development.

## DISCUSSION

### **RAP2 Genes Encode a New Family of Putative DNA Binding Proteins.**

One important conclusion from the characterization of these clones is that the AP2 domain has been evolutionarily conserved in at least *Arabidopsis* and tobacco. In addition, there are two subfamilies of AP2 domain containing proteins in *Arabidopsis* that are designated as the AP2-like and the EREBP-like class of RAP2 proteins. *In vitro* studies have shown that both the EREBP and the AP2 proteins bind to DNA in a sequence specific manner and that the AP2 domain is sufficient to confer EREBP DNA binding activity (Ohme-Takagi, *et al.*, 1995, *Plant Cell* 7, 173-182). From these results and the high degree of sequence similarity between the AP2 domain motifs in AP2, the EREBPs, and the RAP2 proteins, it is concluded that RAP2 proteins function as plant sequence specific DNA binding proteins. Although the exact amino acid residues within the AP2

domain required for DNA binding have not yet been identified, sequence comparisons have revealed two highly conserved motifs referred to as the YRG and RAYD elements within the AP2 domain.

The RAYD element is found in all known AP2 domains and contains a conserved core region that is predicted to form an amphipathic  $\alpha$ -helix (Fig. 4). One hypothesis for the function of this  $\alpha$ -helical structure is that it is involved in DNA binding, perhaps through the interaction of its hydrophobic face with the major groove of DNA (Zubay, G., *et al.*, 1959, *J. Mol. Biol.* 7, 1-20). Alternatively, this structure may mediate protein-protein interactions important for RAP2 functions. These interactions may involve the ability to form homo- or heterodimers similar to that observed for the MADS box family of plant regulatory proteins (Huang, H., *et al.*, 1996, *Plant Cell* 8, 81-94; Riechmann, J. L., *et al.*, 1996, *Proc. Natl. Acad. Sci. USA* 93, 4793-4798) and for the mammalian ATF/CREB family of transcription factors (Hai, T., *et al.*, 1991, *Proc. Natl. Acad. Sci. USA* 88, 3720-3724; O'Shea, E. K., *et al.*, 1992, *Cell* 68, 699-708.).

The conserved YRG element may also function in DNA binding due to the highly basic nature of this region in all RAP2 proteins (Fig. 4). However, the YRG element also contains sequences that are specific for each class of RAP2 protein and may be functionally important for DNA binding. Specifically, the WAAIERD motif is highly conserved in tobacco EREBPs and in EREBP-like RAP2 proteins. By contrast, the WEAR/WESH motif replaces the WAAIERD box in AP2-like RAP2 proteins (Fig. 4). *In vitro* studies suggest that the EREBPs and AP2 recognize distinct DNA sequence elements (Ohme-Takagi, *et al.*, 1995, *Plant Cell* 7, 173-182). It is possible that the WAAIERD and WEAR/WESH motifs may be responsible for DNA binding sequence specificity. The presence of two AP2 domains in AP2 may also contribute to differences in sequence specificity. Although the molecular significance of having one or two AP2 domain motifs is not yet known, genetic and Molecular studies have shown that mutations in either AP2 domain affect AP2 function, implying that both are required for wild-type AP2 activity (Jofuku, K. D., *et al.*, 1994, *Plant Cell* 6, 1211-1225).

In addition to *Arabidopsis* and tobacco, cDNAs that encode diverse AP2 domain-containing proteins have been found in maize, rice, castor bean, and several members of the Brassicaceae including canola (Ohme-Takagi, *et al.*, 1995, *Plant Cell* 7, 173-182; Elliot, R. C., *et al.*, 1996, *Plant Cell* 8, 155-168; Klucher, K M., *et al.*, 1996,

*Plant Cell* 8, 137-153; Wilson, K., *et al.*, 1996, *Plant Cell* 8, 659-671 and Weigel, D., 1995, *Plant Cell* 7, 388-389). This strongly suggests that the AP2 domain is an important and evolutionarily conserved element necessary for the structure or function of these proteins.

**RAP2 Gene Expression in Floral and Vegetative Tissues.** The *AP2*, *RAP2.1*, *RAP2.2*, *RAP2.3*, and *RAP2.4* genes show overlapping patterns of gene expression at the mRNA level in flowers, leaves, inflorescence stems, and roots. However, each gene appears to be differentially regulated in terms of its mRNA prevalence. The overlap in *RAP2* gene activity could affect the genetic analysis of *AP2* and *RAP2* gene functions if these genes are also functionally redundant. For example, in flower development *AP2* and *ANT* show partially overlapping patterns of gene expression at the organ and tissue levels (Jofuku, K. D., *et al.*, 1994, *Plant Cell* 6, 1211-1225; Elliot, R. C., *et al.*, 1996, *Plant Cell* 8, 155-168; Klucher, K M., *et al.*, 1996, *Plant Cell* 8, 137-153; W. Szeto). From single and double mutant analysis it has also been suggested that *AP2* may be partially redundant in function with *ANT* (Elliot, R. C., *et al.*, 1996, *Plant Cell* 8, 155-168). The phenomenon of genetic redundancy and its ability to mask the effects of gene mutation is more clearly demonstrated by the MADS domain containing floral regulatory genes *APETALA1 (API)* and *CAULIFLOWER (CAL)*. Genetic studies have demonstrated that mutations in *cal* show no visible floral phenotype except when in double mutant combination with *ap1* (Bowman, J. L., *et al.*, 1993, *Development Cambridge, U.K.*, 119, 721-743), indicating that *API* is completely redundant in function for *CAL*. The hypothesis that the *RAP2* genes may have genetically redundant functions is supported by the fact that the dominant gain-of-function mutation *tiny* is the only *Arabidopsis* *RAP2* EREBP-like gene mutant isolated to date (Wilson, K., *et al.*, 1996, *Plant Cell* 8, 659-671).

**AP2 Activity Is Detectable in Vegetative Development.** The present analysis of *RAP2* gene expression in wild-type and *ap2-10* plants suggests that *AP2* contributes to the regulation of *RAP2* gene activity throughout *Arabidopsis* development. *RAP2* gene expression is both positively and negatively affected by the absence of *AP2* activity during development. The observed differences in *RAP2.2*, *RAP2.3*, and *RAP2.4* gene expression levels in wild-type and *ap2-10* flowers and vegetative tissues are not apparently due to differences in ecotype because similar changes in gene expression levels

were observed for all three RAP2 genes in stems when ecotype was controlled. The regulation of RAP2 gene expression by *AP2* in stems clearly indicates that unlike other floral homeotic genes *AP2* functions in both reproductive and vegetative development.

5

#### Example 5

This example shows that transgenic plants of the invention bear seed with altered fatty acid content and composition.

10

Antisense transgenic plants were prepared using *AP2*, *RAP2.8*, and *RAP2.1* (two independent plants) using methods described above. The fatty acid content and composition were determined using gas chromatography as described Br  un and Somerville *Plant Physiol.* 113:933-942 (1997). The results are shown in Table VI (for *AP2*) and Table VII (for the *RAP2* genes). As can be seen there, the transgenic plants of the invention have increased fatty acid content as compared to wild-type plants. In addition, the profile of fatty acids is altered in the plants.

Table VI. Analysis of Arabidopsis seed fatty acid content and composition.

	Plants	Seed Mass ( $\mu$ g/seed)	Total Fatty Acid Content ( $\mu$ g/seed)	Fatty acid methylesters (%) (area % by GLC)							
				16:0	18:0	18:1	18:2	18:3	20:1	22:1	24:0
5	<i>Mutant</i>										
	<i>ap2-10</i>	27	3.94	9.4	4.4	2.5	70.8	0.8	6.6	0.1	1.9
10	<i><math>\alpha</math>AP2 transgenic plants</i>										
	C24 15-522 (F1-1)	34	11.0	4.5	3.3	3.1	49.0	3.2	26.3	0.6	4.9
	C34 15-542 (F1-2)	24	2.5	6.3	3.4	3.1	52.1	4.0	28.1	0.4	0.4
	<i>Wild-type</i>										
	C24	20	2.67	9.4	4.9	8.8	62.7	0.7	8.1	0.3	1.9

Table VII. Analysis of Arabidopsis RAP2 antisense transgenic seed fatty acid content and composition.

Plants	Seed Mass (μg/seed)	Total Fatty Acid Content (μg/seed)	Fatty acid methylesters (%) (area % by GLC)							
			16:0	18:0	18:1	18:2	18:3	20:1	22:1	24:0
<i>RAP2 antisense</i>										
<i>transgenic plants</i>										
RAP34-2.8C LE	24	9.7	5.1	3.0	3.6	53.6	3.2	29.3	0.3	0.2
RAP34-2.1A' COL	34	11.0	9.5	1.9	3.0	53.6	4.2	20.3	5.5	0.5
RAP34-2.1D' COL	26	8.1	4.6	2.6	2.6	50.3	3.2	33.2	0.3	1.0
<i>Wild-type</i>										
COL*	24	ND	6.0	4.0	14.0	27.0	18.0	22.0	2.0	ND

ND, not determined

\*Patak et al. (1994) Oil content and fatty acid composition of seeds of various ecotypes of *Arabidopsis thaliana*: a search for useful genetic variants. *Curr. Sci.* 67, 470-472.



## Example 6

This example describes construction of promoter construct which are used to prepare expression cassettes useful in making transgenic plants of the invention. In particular, this example shows use of two preferred promoters, the promoter from the *AP2* gene and the promoter from the *Bell* gene.

Figure 5 shows a *AP2* promoter construct. pAP2 represents the 16.3 kb *AP2* promoter vector cassette that is used to generate chimeric genes for use in plant transformations described here. pAP2 is comprised of the 4.0 kb promoter region of the *Arabidopsis* *AP2* gene. The Ti plasmid vector used is pDE1000 vector (Plant Genetic Systems, Ghent, Belgium). The pDE1000 vector DNA was linearized with BamH1 and the *AP2* promoter region inserted as a 4.0 kb BamH1 DNA fragment from plasmid subclone pLE7.2. At the 3' end of the inserted *AP2* promoter region, designated *AP2*, lie three restriction sites (EcoR1, Sma1 and SnaB1) into which different gene coding regions can be inserted to generate chimeric *AP2* promoter/gene cassettes. NOS::NPTII represents the plant selectable marker gene NPTII under the direction of the nopaline synthase promoter which confers resistance to the antibiotic kanamycin to transformed plants cells carrying an integrated *AP2* promoter cassette. LB and RB represent the T-DNA left and right border sequences, respectively, that are required for transfer of T-DNA containing the *AP2* promoter cassette into the plant genome. PVS1 designates the bacterial DNA sequences that function as a bacterial origin of replication in both *E. coli* and *Agrobacterium tumefaciens*, thus allowing pAP2 plasmid replication and retention in both bacteria. Amp<sup>R</sup> and Sm/Sp<sup>R</sup> designate bacterial selectable marker genes that confer resistance to the antibiotics ampicillin and streptomycin/spectinomycin, respectively, and allows for selection of *Agrobacterium* strains that carry the pAP2 recombinant plasmid.

Figure 6 shows a *BEL1* promoter construct. pBEL1 represents the 16.8 kb *BEL1* promoter vector cassette that is used to generate chimeric genes for use in plant transformations described here. pBEL1 is comprised of the 4.5 kb promoter region of the *Arabidopsis* *BEL1* gene. The Ti plasmid vector used is pDE1000 vector (Plant Genetic Systems, Ghent, Belgium). The pDE1000 vector DNA was linearized with BamH1 and the *BEL1* promoter region inserted as a 4.5 kb BamH1-Bgl2 DNA fragment from plasmid subclone pλ1C9R (L. Reiser, unpublished). At the 3' end of the inserted *BEL1* promoter region, designated *BEL1*, lie three restriction sites (EcoR1, Sma1 and SnaB1) into which

different gene coding regions can be inserted to generate chimeric BEL promoter/gene cassettes. NOS::NPTII represents the plant selectable marker gene NPTII under the direction of the nopaline synthase promoter which confers resistance to the antibiotic kanamycin to transformed plants cells carrying an integrated BEL promoter cassette. LB and RB represent the T-DNA left and right border sequences, respectively, that are required for transfer of T-DNA containing the BEL promoter cassette into the plant genome. PVS1 designates the bacterial DNA sequences that function as a bacterial origin of replication in both *E. coli* and *Agrobacterium tumefaciens*, thus allowing pBEL plasmid replication and retention in both bacteria. Amp<sup>R</sup> and Sm/Sp<sup>R</sup> designate bacterial selectable marker genes that confer resistance to the antibiotics ampicillin and streptomycin/spectinomycin, respectively, and allows for selection of *Agrobacterium* strains that carry the pBEL recombinant plasmid.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference for all purposes.

## SEQUENCE LISTING

## SEQ ID NO:1 Canola ADC gene

Insert size: 559 bp

Location of primer RISZU1 nt 1-21

Location of primer RISZU2 nt 544-559

PREDICTED 5'/3' ENDS OF:

exon 2 nt 1-26

intron 2 nt 27-169

exon 3 nt 170-200

intron 3 nt 201-306

exon 4 nt 307-394

intron 4 nt 395-473

exon 5 nt 474-559 NOTE: RISZU2 ends within coding region for exon 5

5'-GGACTGTGGGAAACAAGTTTACTTAGGTATGATCATGTAATGTTGTTCAAACACAGATCAAATAT  
CCTATTGA

AACTAAGTTGTGTTGTGTCCGTCCATTTTTATATGATTTCTTCGACCAAATAAAGGTTTNANTATCTC  
CTTATATT

ACTTTTGTACATATTCAGGTGGATTTGACACAGCACATGCCGCTGCTCGGTATGTTTAACTCATCC

AAATATGA

TCAATTAGAACGAATCTAATATTCCTTATTTGTAATTTGCTGATATACAAATTAATTTGGGTGGGTAA  
CTGTTTGG

GACAGTGCCTACGATAGAGCCGCAGTTAAGTTTAGAGGTGTNATGCAGAATATNANTTNCAANATT  
GAAGACTANG

TGGAGGANTTGAAACAGGTAGAACATCTATCATTGGTNGGNCCAAGCNNTNGACCNANATCTTAC  
TCCACCNTCC

NCCTNTAATNCNTGNTGCNGATGAGCACCTTGACAANGGACGAGTCATGCNTGTCATTAGNCGCNA  
NCCCTGGGTT

CCANNAAGCNCCTCAAGTANAGANGGTCAC-3'

## SEQ ID NO:2. SOYBEAN ADC GENE

Insert size: ~800 bp

a. Partial sequence generated by sequencing from the 5' end of gene nt 1-291

Location of primer RISZU1 nt 1-21

PREDICTED 5'/3' ENDS OF:

exon 2 nt 1-27

intron 2 nt 28-264

exon 3 nt 265-291

(NOTE: partial sequence; coding region for exon 5 extends beyond the sequence given here)

5'-GGATTGTGGGAAACAAGTTTATCTAGGTAAAGTTGAYTAAYAACAATAATTGTATATGTGTTTGTG  
AGAACTG  
5 TGGCASTTATTTTCCCTAATTATCGTTTAAAGACGCTAAMACGGTTTTTTTTCCTTGTCTTGTGTTTT  
TTGYCTT  
GGCTGTGAYGCGGTAAAMACAANAGTGTGAGTGTGTGTTGTGTGTGGGTGAGGAYTYTTCCINTC  
NCCGTGGTGA  
10 CTGACTTGATGGNTCTTGCTGGGTNCANNTTNTCTACGTGGATTCGACACNNCACATGCGGCNGC-  
3'

b. Partial sequence generated by sequencing from the 3' end of gene nt 1-972

Location of primer RISZU2 nt 952-972

PREDICTED 5'/3' ENDS OF:

exon 4 nt 629-716

15 intron 5 nt 717-876

exon 5 nt 877-end

(NOTE: Sequence below represents the reverse/complement of the sequence  
generated and therefore runs from the 5' to the 3' end of the soybean AP2  
gene sequence. RISZU2 ends within coding region for exon 5.)

20 5'-NCGGGACGGACNGGNCGGANGCNANGCGCNNNNAGACGGNANCGCAANNACGNCNCNTCGCCT  
CCGNGGGCCC  
NACNCNCTCGGACGGNTGNGAGGNCNCCNNGNNTCNGCNGGNAGCANGNGGTGCNGCCNTGG  
CGANCCGCCCCG  
25 NNGAGNGNAGNCNTGGNGNCGACACCNTGCAGNCNTCTNGCNATGNGATGGNNAACNGGAGACCG  
GAGTGACNGTC  
GNGAGNAANGCGANANNTNNTGGTGNTCCCGGGCGNGAANGNCCNGACAGATGGGTGGAACGTAT  
TGCAAGTGANA  
CCAAGAAGGGTNGGACGGCGTATTTCTAATGTTAGGNTANNNTTNTCCTTTGGTTANTCTGCCNGN  
30 GCGAAACNG  
GGGAGATGGNNNGGGGNGAGATTTTTNTNTGNGNGACGACTGANGNNTCTGTTGGGTNAATTGTCT  
AGGGGATTGA  
CACNCACATGCGTGCTCGTGGCCCTGCCCTTCCTTCAGTATNATACCCAAGCTTGATNTTACTTT  
NTCCATGTC  
35 TTGAACCAAATATCAAATATTATTGTNAATCACATTCGTTGTGGNCCGGGAATTGTGAGTCTCAA  
GAAAATTGT  
GTATTNTCCGTCTCTCTTTTCAGTGCTTATGATAGAGCGGCTATTAAATCCGAGGAGTGGAGGCTG  
ACATTAAC

TCAATATTGNAGACTATGAAGATGACTTGAAGCAGGTGATCAATTGTGGATTATGTTTTTTTTATT  
GAATAAAT  
GCATTTATCGTATTTATCTTATCTTACAGTCATACGTATAGGATGCACCTTATCTCCCACAGTTAGTG  
TTTTTTTT  
5 ATCTGAATTATTCTCATGATTTTGTTAAATGCAATGTTAATAGATGAGCAATCTTACCAAGGAAGAG  
TTCGTCCAC  
GTGCTTCGCCGCCAAAGCACTGGATTTCCGAGAGGAAGCTCCAAGTATAGAGGTCACTTTGCA-3',

Designations for sequence ambiguities

10

N = A,G,C,T

Y = C,T

S = G,C

M = A,C

15

WHAT IS CLAIMED IS:

1. A method of modulating seed mass in a plant, the method comprising:  
providing a first plant comprising a recombinant expression cassette containing  
an *ADC* nucleic acid linked to a plant promoter;  
selfing the first plant or crossing the first plant with a second plant, thereby  
producing a plurality of seeds; and  
selecting seed with altered mass.
2. The method of claim 1, wherein expression of the *ADC* nucleic acid  
inhibits expression of an endogenous *ADC* gene and the step of selecting includes the step  
of selecting seed with increased mass.
3. The method of claim 2, wherein the seed have increased protein  
content, carbohydrate content, or oil content.
4. The method of claim 2, wherein the *ADC* nucleic acid is linked to the  
plant promoter in the antisense orientation.
5. The method of claim 2, wherein the *ADC* nucleic acid hybridizes under  
stringent conditions to a nucleic acid having a sequence as set forth in Genbank Accession  
Nos. U12546, AF003101, and AF003094.
6. The method of claim 2, wherein the *ADC* nucleic acid hybridizes under  
stringent conditions to a nucleic acid having a sequence as set forth in SEQ ID NO:1.
7. The method of claim 2, wherein the *ADC* nucleic acid hybridizes under  
stringent conditions to a nucleic acid having a sequence as set forth in SEQ ID NO:2.
8. The method of claim 2, wherein the *ADC* nucleic acid is selected from  
a group consisting of Genbank accession numbers U12546, AF003094, AF003095,

AF003096, AF003097, AF003098, AF003099, AF003100, AF003101, AF003102, AF003103, AF003104, and AF003105.

5           9.     The method of claim 2, wherein the first and second plants are the same species.

          10.    The method of claim 2, wherein the first and second plants are members of the family Brassicaceae.

10           11.   The method of claim 2, wherein the first and second plants are members of the family Solanaceae.

          12.    The method of claim 2, wherein the plant promoter is a constitutive promoter.

15           13.    The method of claim 12, wherein the promoter is a CaMV 35S promoter.

          14.    The method of claim 2, wherein the promoter is a tissue-specific promoter.

20           15.    The method of claim 14, wherein the promoter is ovule-specific.

          16.    A seed produced by the method of claim 2.

25           17.    The method of claim 1, wherein expression of the *ADC* nucleic acid enhances expression of an endogenous *ADC* gene and the step of selecting includes the step of selecting seed with decreased mass.

          18.    The method of claim 17, wherein the *ADC* nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in Genbank Accession Nos. U12546, AF003101, and AF003094.

30           19.    The method of claim 18, wherein the *ADC* nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in Genbank Accession Nos. U12546, AF003101, and AF003094.

19. The method of claim 17, wherein the *ADC* nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in SEQ ID NO:1.

5 20. The method of claim 17, wherein the *ADC* nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in SEQ ID NO:2.

10 21. The method of claim 17, wherein the *ADC* nucleic acid is selected from a group consisting of Genbank accession numbers U12546, AF003094, AF003095, AF003096, AF003097, AF003098, AF003099, AF003100, AF003101, AF003102, AF003103, AF003104, and AF003105.

15 22. The method of claim 17, wherein the first and second plants are the same species.

23. The method of claim 17, wherein the first and second plants are members of the family Brassicaceae.

20 24. The method of claim 17, wherein the first and second plants are members of the family Solanaceae.

25 25. The method of claim 17, wherein the plant promoter is a constitutive promoter.

26. The method of claim 25, wherein the promoter is a CaMV 35S promoter.

30 27. The method of claim 17, wherein the promoter is a tissue-specific promoter.

28. The method of claim 27, wherein the promoter is ovule-specific.



29. A seed produced by the method of claim 17.

30. A seed comprising a recombinant expression cassette containing an ADC nucleic acid.

31. The seed of claim 30, which is derived from a plant that is a member of the family Brassicaceae.

32. The seed of claim 30, wherein the ADC nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in Genbank Accession Nos. U12546, AF003101, and AF003094.

33. The seed of claim 30, wherein the ADC nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in SEQ ID NO:1.

34. The seed of claim 30, wherein the ADC nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in SEQ ID NO:2.

35. The seed of claim 30, wherein the ADC nucleic acid is selected from a group consisting of Genbank accession numbers U12546, AF003094, AF003095, AF003096, AF003097, AF003098, AF003099, AF003100, AF003101, AF003102, AF003103, AF003104, and AF003105.

36. The seed of claim 30, wherein the ADC nucleic acid is linked to a plant promoter in an antisense orientation and the seed mass is at least about 10% greater than the average mass of seeds from the same plant variety which lack the recombinant expression cassette.

37. The seed of claim 36, wherein the mass is at least about 20% greater than the average mass of seeds from the same plant variety which lack the recombinant expression cassette.

38. The seed of claim 36, wherein the mass is at least about 50% greater than the average mass of seeds from the same plant variety which lack the recombinant expression cassette.

5 39. The seed of claim 36, wherein the oil content is proportionally increased.

40. The seed of claim 36, wherein the protein content is proportionally increased.

10 41. The seed of claim 30, wherein the *ADC* nucleic acid is linked to a plant promoter in the sense orientation and the seed mass is at least about 10% less than the average mass of seeds of the same plant variety which lack the recombinant expression cassette.

15 42. The seed of claim 41, which has a mass at least about 20% less than the average mass of seeds of the same plant variety which lack the recombinant expression cassette.

20 43. The seed of claim 41, which has a mass at least about 50% less than the average mass of seeds of the same plant variety which lack the recombinant expression cassette.

25 44. A transgenic plant comprising an expression cassette containing a plant promoter operably linked to a heterologous *ADC* polynucleotide.

45. The transgenic plant of claim 44, wherein the *ADC* nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in Genbank Accession Nos. U12546, AF003101, and AF003094.

46. The transgenic plant of claim 44, wherein the *ADC* nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in SEQ ID NO:1.

5 47. The transgenic plant of claim 44, wherein the *ADC* nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in SEQ ID NO:2.

10 48. The transgenic plant of claim 44, wherein the *ADC* polynucleotide is selected from a group consisting of Genbank accession numbers U12546, AF003094, AF003095, AF003096, AF003097, AF003098, AF003099, AF003100, AF003101, AF003102, AF003103, AF003104, and AF003105.

15 49. The transgenic plant of claim 44, wherein the heterologous *ADC* polynucleotide encodes a *ADC* polypeptide.

50. The transgenic plant of claim 44, wherein the heterologous *ADC* polynucleotide is linked to the promoter in an antisense orientation.

20 51. The transgenic plant of claim 44, which is a member of the genus *Brassica*.

25 52. An isolated nucleic acid molecule comprising an expression cassette containing a plant promoter operably linked to a heterologous *ADC* polynucleotide.

53. The isolated nucleic acid molecule of claim 52, wherein the *ADC* nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in Genbank Accession Nos. U12546, AF003101, and AF003094.

30 54. The isolated nucleic acid molecule of claim 52, wherein the *ADC* nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in SEQ ID NO:1.

55. The isolated nucleic acid molecule of claim 52, wherein the *ADC* nucleic acid hybridizes under stringent conditions to a nucleic acid having a sequence as set forth in SEQ ID NO:2.

5 56. The isolated nucleic acid molecule of claim 52, wherein the *ADC* polynucleotide is selected from a group consisting of Genbank accession numbers U12546, AF003094, AF003095, AF003096, AF003097, AF003098, AF003099, AF003100, AF003101, AF003102, AF003103, AF003104, and AF003105.

10 57. The isolated nucleic acid of claim 52, wherein the heterologous *ADC* polynucleotide encodes a *ADC* polypeptide.

58. The isolated nucleic acid of claim 52, wherein the heterologous *ADC* polynucleotide is linked to the promoter in an antisense orientation.

15 59. The isolated nucleic acid of claim 52, which is a member of the genus *Brassica*.

195  
 AP2-R1 SSQYRGVTFYRRTGRWESH I --WDCGKQVYLGGFDTAHAAARAYDRAAIKFRGVEADINFNIDDYDDDL  
 221  
 AP2-R2 SSKYRGVTLHKC-GRWEARMGQFLGKKYVYLGLFDTEVEAARAYDKAAIKCNGKDAVTNFOPSIYDEEL  
 288  
 op2-5V  
 op2-1  
 op2-10

FIG. 1A.

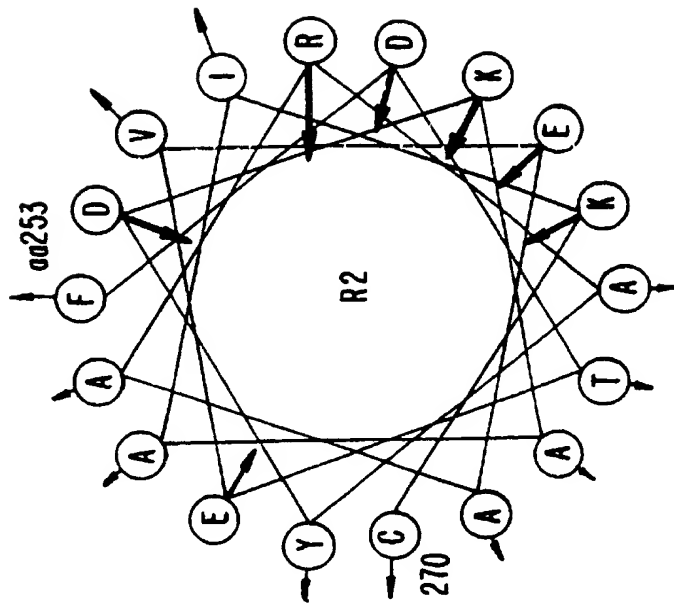
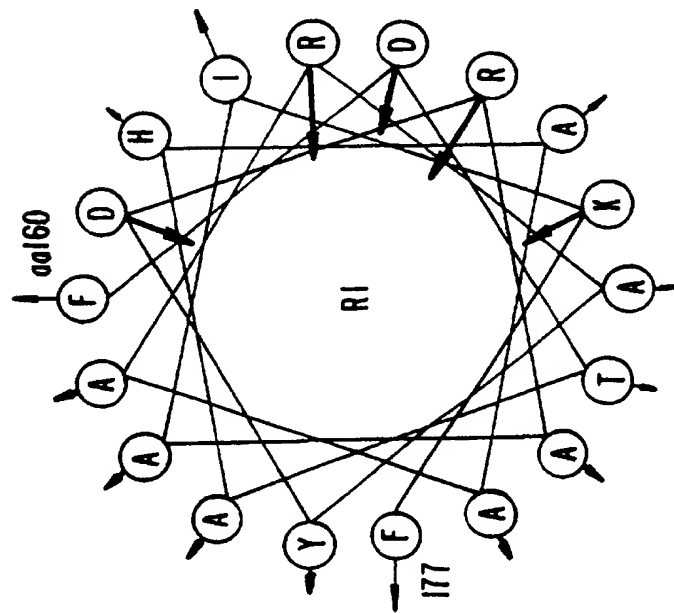
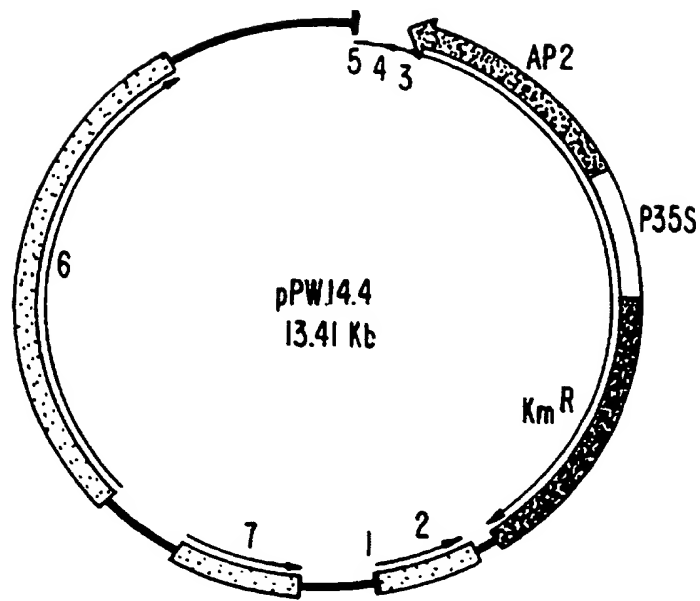
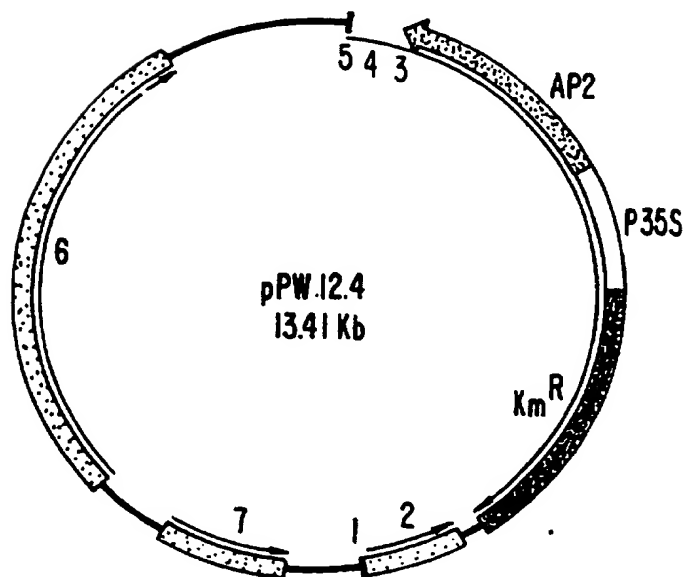


FIG. 1B.

2/7

**FIG. 2.****FIG. 3.**

SUBSTITUTE SHEET (RULE 26)

	YRG ELEMENT				RAYD ELEMENT				
					α-HELIX				
	1	16	31	46	61	76			
AP2-R1	SSQYRQVT	WESHWD	--CGKQ	DTA	HAAARAYDRAAAIKER	GVEADINENIDDD	DD		
-R2	SSKYRGVT	WEARMQFLG	--KYYQYLG	DTE	VEAARAYDKAAIKCN	GKDAVTNFDPSI	DE		
ANT-R1	TSQYRGVTR	YEAHLWDNSFKKEGH	SRKGFQVYDGGYDME	EKAARAYDILAAALKYW	GPSHTHNFSAENYQK				
-R2	ASIFRFVTR	WQARIGRVAG	--NKDLYLQ	GTQ	EEAAPAYDMAATKER	GTNAVTFNFDITRYDV	CR		
RAP2.7-R1	SSQYRGVTFYRRRTGR	WASHIWD	--CGKQYVLG	DTA	HAAARAYDRAAAIEFR	GVDADCNFTLGDYEE	DM		
-R2	SSKYRGVT	WEARM	--QOE	LGK	---	GREAVTNFEMSSYQN	EI		
AP2-R1	67								
-R2	68								
ANT-R1	77								
-R2	69								
RAP2.7-R1	67								
-R2	53								

```

wear
CONSENSUS .S.YRGVT.....GR wesh.....+ .vYLG.f... ..AA.AYD.AAik.. G....tNF.....Y... -.
i

```

**FIG. 4A.**

**SUBSTITUTE SHEET (RULE 26)**

	YRG ELEMENT					16		RAYD ELEMENT					46	61	761		
	++	+	+	++	+			---	---	---	---	EEAAIAYIK	KAAYRMF	CSKKHILN	FFPLEVANT	KQ69	
EREBP-1	GRHYRGVR	-RRPWGM				FAAEIRDR	PAKNG	---	---	ARVWLGT	YETD	EEAAIAYIK	KAAYRMF	CSKKHILN	FFPLEVANT	KQ69	
EREBP-2	GRHYRGVR	-QRPWGM				FAAEIRDR	PAKNG	---	---	ARVWLGT	YETA	EEAALAYIK	KAAYRMF	CSKALLP	FFPHRIGLN	EP69	
EREBP-3	EVHYRGVR	-KRPWGR				YAAEIRDR	PKKS	---	---	RVWLGT	FTDA	EEAAKAYD	TAAAREFP	CPKAKTN	FFSPTENQ	SP68	
EREBP-4	KKHYRGVR	-QRPWGR				FAAEIRDR	NRKG	---	---	TRVWLGT	PTDA	IFAAKAYD	RAAEKLR	GSKAIVN	FFPHRIGLN	EP69	
	+	+	+	+	+		++										
RAP2.2	KNQYRGIR	-QRPNGA				NAAEIRDR	RXGS	---	---	REWLGTP	DTA	NEAARAYD	AAARRIR	QTKAKVN	EEKNPS	VV64	
RAP2.3	KNVYRGTR	-KRPWGR				WAAEIRDR	RKGV	---	---	RVWLGT	FTNTA	EEAAMAYD	VAAKQIR	QOKAKIN	FFDLHPP	PP68	
RAP2.5	EIRYRGVR	-KRPWGF				YAAKIRDR	PKKT	---	---	RVWLGT	FTDA	KKAARAYD	TAAARDEF	CAKAKTN	FFTFLELS	DQ68	
RAP2.6	PKKYRQVR	-QRPWGM				WAAIRDR	PHKAT	---	---	RVWLQTF	ETDA	TAAATAYD	AAALREF	GSKAKIN	FFENVGTQ	TI68	
RAP2.12	KNQYRGIR	-QRPWGR				WAAVIRDR	REGA	---	---	RIWLGT	FTKTA	FEAARAYD	AAARRIE	GSKAKVN	FFREENMKA	NS68	
	+	+	++	+	+		++										
TINY	HPVYRGVR	-KRPWGM				WVSEIREP	RKKS	---	---	RIWLGT	FPHP	EMAARAH	DVAALLSIR	GASWILN	FFPOLAGSF	PR68	
RAP2.1	RKPYRGIR	-RRKWGK				WVARIRKR	NKRS	---	---	RLMLGSY	TTD	IAAARAYD	VAVEYLR	CP SARIN	FFESLLIQE	ED68	
RAP2.4	TKLYRGVR	-QRHWGK				WVAKIRLR	NRNT	---	---	RLMLGT	FTDA	ELAAALAYD	KAAYKLF	GDFARIN	FFENLRHNG	FH68	
RAP2.8	SSKYKGVV	-PPPNGK				WGAQIYKH	Q	---	---	KVWLGT	FTNED	ELAARSYD	IACCRFF	CRDAVVN	FFKNVLEDG	DL66	
RAP2.10	DKPYKQIR	-MRKWGK				WVARIRKR	NKRS	---	---	RIWLGSY	STP	EAAARAYD	TAVFYLE	CP SARIN	FFPELLAGV	TV68	
RAP2.11	KTKKVGVR	-QRP SGK				WVAKIKDT	TQKI	---	---	QMWLGTF	ETDA	KEAARAYD	EAACLLR	GSNTRTN	FFANHFPPN	SQ68	
						fv	ke			sy		i	v				
CONSENSUS...	YRGvr..	R.wG+				waAEIRD	.....			R.WLGt	f.t.	eeAA+AYD	Aa...				G..A..NFP
						Y						I					..

**FIG. 4B.**

Sequence	Position
R1 -KQMTNLTREEFVHVLRRLRQSTGFGRG-	R2
R1 -EDMMKNMTRQEVVAHLRLRKSSGFSGR-	R2
R1 -MKQVQNLSKEEFVHILRLRQSTGFSGR-	R2
R1 -..qm.Nlt+eefvh..LRRqstGFSGR-	R2
CONSENSUS	

**FIG. 4E.**



5/7

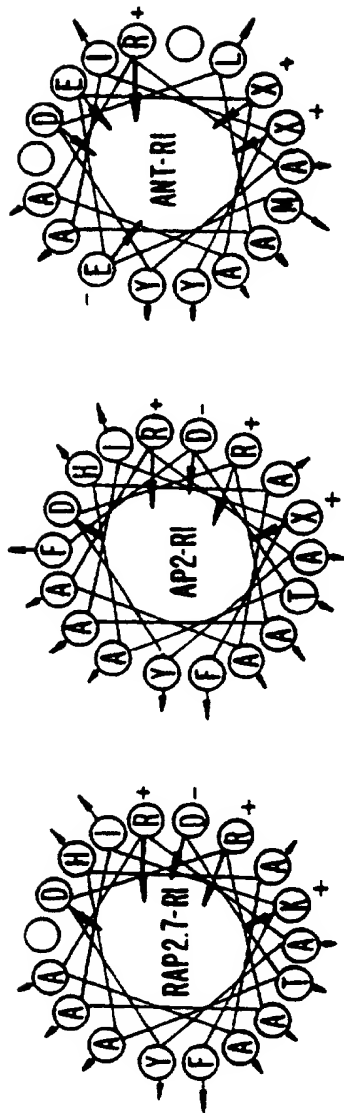


FIG. 4C.

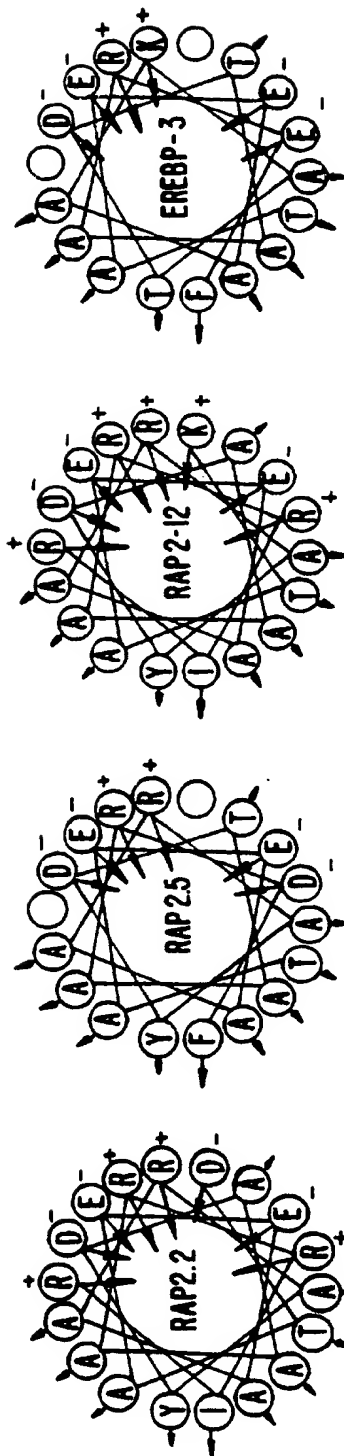
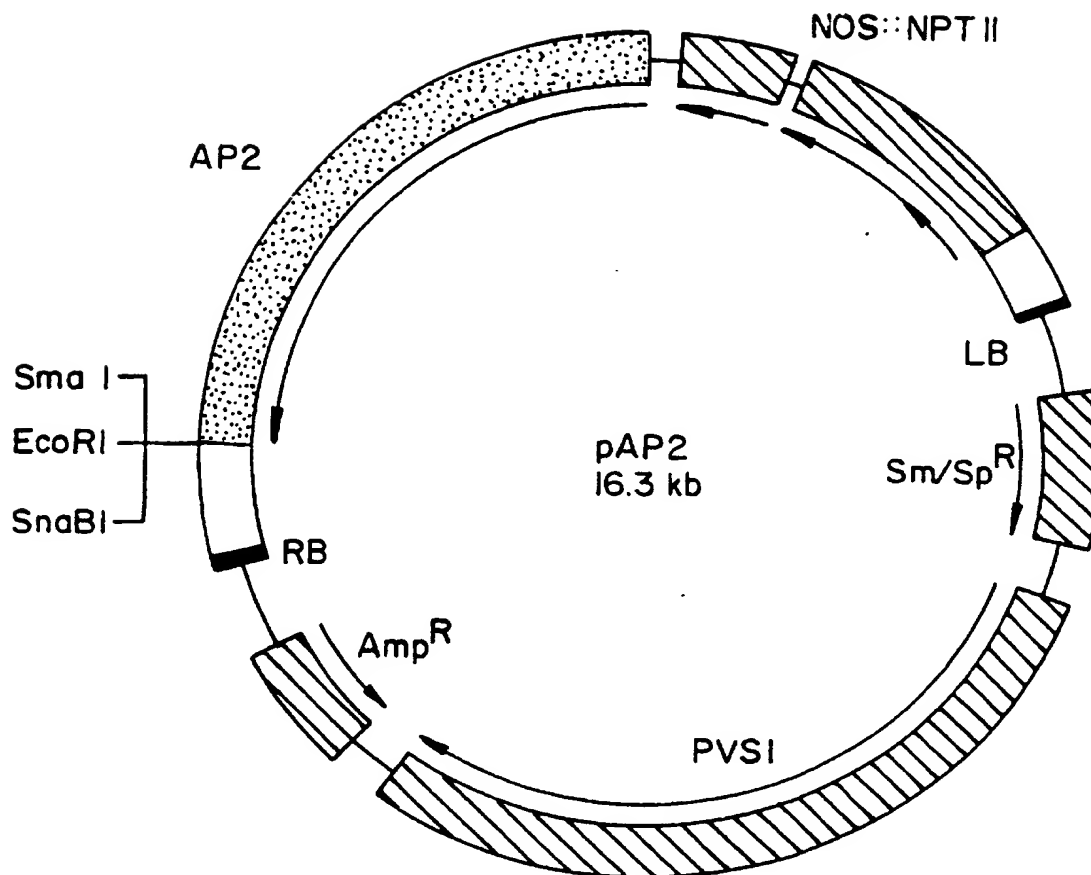


FIG. 4D.

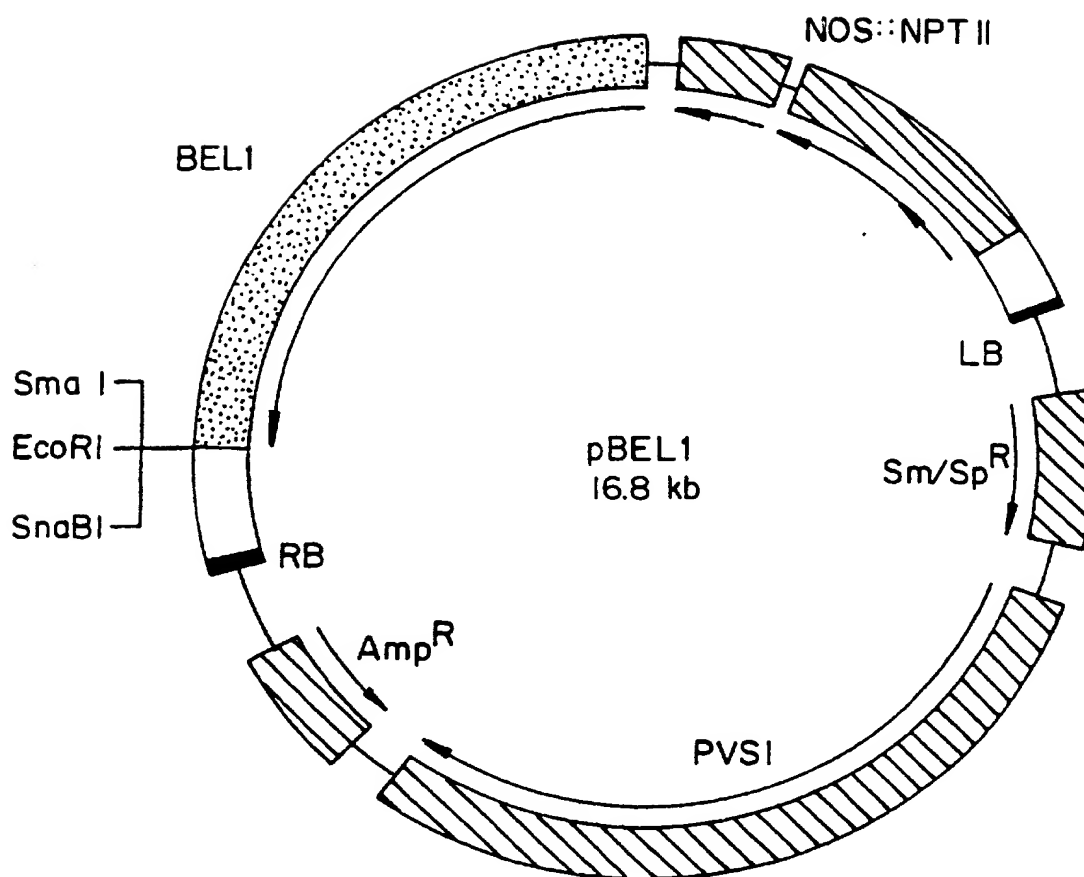
SUBSTITUTE SHEET (RULE 26)

6/7

**FIG. 5.**

SUBSTITUTE SHEET (RULE 26)

7/7

**FIG. 6.**

SUBSTITUTE SHEET (RULE 26)

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/14659

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : C12N 15/00, 15/29; A01H 1/00, 5/00, 5/10

US CL : 435/172.3; 800/205, 250, 255; 536/23.6; 47/58

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/172.3; 800/205, 250, 255; 536/23.6; 47/58

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, BIOSIS, DERWENT WORLD PATENTS. SEARCH TERMS: APETALA?, AP2, AP 2, SEED, SEEDS, SIZE, MASS, WEIGHT, TRANSGENIC?

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- A	JOFUKU et al. Control of Arabidopsis Flower and Seed Development by the Homeotic Gene APETALA2. The Plant Cell. Vol. 6. pages 1211-1225, especially page 1215, column 1, lines 3-7.	44, 45, 48-53, 56-59 ----- 1-5, 8-18, 21-32, 35-43

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* "A"	Special categories of cited documents: document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E"	earlier document published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means	"A" document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

22 SEPTEMBER 1997

Date of mailing of the international search report

31 OCT 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

MARY E. MOSHER, PH.D.

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)\*

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/14659

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☒ Claims Nos.: 6, 7, 19, 20, 33, 34, 46, 47, 54, 55  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  
  
NO COMPUTER READABLE FORM OF SEQ ID NO:1 OR SEQ ID NO:2 WAS PROVIDED.
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1))(July 1992)\*

